

Term Structure Models with Unspanned Factors and Unspanned Stochastic Volatility

Alex Backwell



July 2018

*Thesis Presented for the Degree of
Doctor of Philosophy
in the Department of*

The African Institute of Financial Markets and Risk Management

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Abstract

Certain models of the term structure of interest rates exhibit *unspanned stochastic volatility* (USV). A model has this property if it involves a source of stochastic variation — called an *unspanned factor* — that does not affect the model’s interest rates directly, but does affect the extent to which future interests are liable to change (that is, interest-rate *volatility*). This thesis is concerned with these models, from a variety of perspectives.

Firstly, the theoretical foundation of the USV property is addressed. Formal definitions of unspanned factors and USV are developed, generalising ones tentatively proposed in the literature. Several results from these definitions and the accompanying framework are derived. Particularly, the ability to hedge general claims (i.e., the *completeness* or lack thereof) of these models is examined in detail. Examples are given to illustrate the features of the proposed framework and the necessity of the generalised definitions.

Secondly, the empirical issue of whether USV models are necessary to plausibly represent observed interest-rate markets is interrogated. An empirical derivative-hedging approach is adopted, the results of which are contextualised by also treating data simulated from models with USV and non-USV versions. It is shown that hedging effectiveness is relatively robust to the presence of USV, which resolves the apparent conflict between the two studies that have taken a hedging approach to this question. Despite the cross-sectional hedging effects being surprisingly minor, further regression results show that USV models are needed to model the time series of market interest rates.

Finally, the thesis addresses a certain class of models that exhibit USV: those with one spanned factor (driving interest-rate variation) and one unspanned, volatility-related factor. Being the simplest non-trivial USV models, these *bivariate USV models* are fundamental, and — like one-factor models in general settings — are helpful in introducing and comparing higher-factor models when simple ones are insufficient. These models are shown to exist (contradicting a claim in the literature); to share a particular affine form for their bond pricing functions; and to necessarily exhibit a short-term interest rate with dynamics of a certain type. A specific bivariate USV model is then proposed, which is analysed and compared to others in the literature.

Acknowledgements

The author acknowledges and is extremely thankful for funding assistance from the University of Cape Town and the National Research Foundation of South Africa.

The attentive and insightful supervision of Peter Ouwehand deserves recognition and is greatly appreciated. The thesis would not have been possible without David Taylor, whose role has been indirect but considerable. Thanks also go out to Ralph Rudd.

Thanks are due to the following individuals, who have provided valuable advice, mentorship and assistance over the last several years: Thomas McWalter, Andrea Macrina, Martin Larsson, Melusi Mavuso, Jörg Kienitz, Obeid Mahomed, Maria Kulikova, Stéphane Crépey, Peter Ritchken, Eric Schlögl, Dave Strugnell, Coenraad Labuschagne, Dorje Brody and Lameez Fakier.

Finally, the debt of gratitude owed to Jessica Fell and to Karin and Peter Backwell cannot be overstated. Thank you.

Contents

1	Introduction	1
1.1	Interest-rate and Term Structure Modelling	1
1.2	Unspanned Stochastic Volatility	4
1.3	Thesis and Contribution Overview	8
1.3.1	Chapter 2	8
1.3.2	Chapter 3	9
1.3.3	Chapter 4	11
1.3.4	Chapter 5	12
2	Mathematical Introduction and Foundation	14
2.1	Formal Framework	14
2.2	Additional Definitions	18
3	Unspanned Factor Theory	24
3.1	Unspanned Factors	25
3.2	Unspanned Factor Roles	41
3.3	Proofs of Completeness Results	48
3.3.1	Proof of Theorem 3.1.9	48
3.3.2	Proof of Theorem 3.2.6	51
3.3.3	Proof of Theorem 3.2.10	54
3.4	Ensuing Remarks and Examples	65
3.4.1	Partially Spanned Factors	65
3.4.2	Time-inhomogeneous Models	66
3.4.3	Non-Brownian Cases	68
3.4.4	Unspanned Stochastic Correlation	68
4	Empirical USV Evidence	70
4.1	Empirical Approach	70
4.1.1	Evidence Review	70
4.1.2	Data and Methodology	74
4.2	Results	81
4.2.1	Original USV Evidence	81
4.2.2	Basic Hedging Results	84
4.2.3	Hedging and USV	85
4.2.4	Secondary Tests	86
4.2.5	The Two-tier Test	88
4.3	Conclusion	90

5	Bivariate USV Models	92
5.1	(1,1) Models	94
5.2	The LADQ(1,1) Model	100
5.2.1	Risk-neutral Specification	100
5.2.2	Comparison to Other Models	102
5.2.3	LADQ(1,1) Term Structure	104
5.2.4	LADQ(1,1) Option Pricing	104
5.2.5	LADQ(1,1) Time-series Dynamics	106
5.2.6	Yield Curve Calibration	108
5.2.7	Empirical Calibration Example	109
5.3	Extensions and Conclusion	110
5.3.1	Non-zero Lower Bounds	110
5.3.2	Further Volatility State Variables	112
5.3.3	One-factor Models	112
A	Existence of Suitable State Process Transformations	113
B	Invertible Matrix Processes	117
C	Partial Derivatives of the Bond Pricing Function	119
C.1	The Short Rate and the Bond Pricing Function	119
C.2	Sensitivity to the Short Rate	120
D	Comparison to Linear-rational Models	121
D.1	(1,1) Models in the Linear-rational Framework	121
D.2	The Short Rate in the LRSQ(1,1)Model	123

Chapter 1

Introduction

The thesis begins, in Section 1.1, with a brief review of term structure modelling. Section 1.2 then gives an introduction to unspanned stochastic volatility, after which an overview of the thesis and its contributions is given in Section 1.3.

1.1 Interest-rate and Term Structure Modelling

Uncertainty around future interest rates is a significant source of financial risk. Representing this uncertainty with a mathematical model can be extremely helpful in managing this risk. For example, one can follow Vasiček [1977] by supposing that the short-term interest rate over time is a continuous-time stochastic process $\{r_t\}$ satisfying

$$dr_t = \kappa(\theta - r_t) dt + \sigma dW_t, \quad (1.1)$$

where κ , θ and σ are constants and where $\{W_t\}$ is a (standard, one-dimensional) Brownian motion. The exact nature of the *short rate* r_t is clarified in Chapter 2, as are the key formalities and technical details of the thesis (such as the probability space on which $\{W_t\}$ is defined). It suffices for now to note that a certain asset available for trading — namely, the *cash account* — offers the short rate as its annualised return over each instant.¹ Letting $\{B_t\}$ denote the cash account value process, it follows that

$$B_t = B_0 e^{\int_0^t r_u du}. \quad (1.2)$$

Note that the short rate is the cash account's (annualised, instantaneous) return, not merely its expected return — this makes the cash account risk free in an important sense clarified in Chapter 2.

Having found suitable values for the parameters κ , θ and σ (by, for example, statistically inferring them from a history of short rate realisations), Equation 1.1 implies a probability distribution for the future short rate — the Vasiček [1977] specification is thus an *interest-rate model*. Such a representation of interest-rate uncertainty is instrumental in financial risk management. One can, for instance, use the model to calculate the probability of the future short rate exceeding a certain level. Many financial positions depend on short-term interest rates; one can calculate or estimate these dependencies and use an interest-rate model to represent the uncertainty in a variety of positions and portfolios. The Brownian motion in the Vasiček [1977] or some other

¹In our treatment of interest-rate modelling, the cash account is unique — the absence of tenor or credit spreads is due to the *pre-crisis* nature of the framework that Chapter 2 establishes.

Brownian-based model could be correlated to Brownian motions used in the modelling of other financial risks, allowing, for example, the *value at risk* of an investment portfolio to be estimated in a way that accounts for interest-rate risk.²

The plausibility of a financial model — the extent to which the model accurately represents the relevant uncertainties — is clearly a crucial determinant of its usefulness and applicability. For a financial model to be plausible, its implied distributions must conform to established statistical properties of the relevant variables. It must also preclude *arbitrage*; that is, it must not allow risk-free, cost-free trading profits to be made, as these are absent (or at least extremely fleeting) in liquid markets [Piazzesi, 2010, §1.3]. Of course, other factors besides plausibility influence a model’s usefulness (such as its tractability, which is often in tension with its plausibility). In a very general sense, this thesis explores the plausibility of different classes of interest-rate models.

A central concept in financial modelling, including term structure modelling, is that of a *risk-neutral measure*: the probabilities of events in a model might be adjustable (that is, there may exist a different probability measure) in a way that causes the ratio of any tradeable asset price process and the cash account process to be a *martingale* under the adjusted probabilities.³ Critically, the existence of such a risk-neutral measure is a sufficient condition for the model in question to be free of arbitrage.⁴ It is conventional to assume that a risk-neutral measure exists, and to thereby ensure that the model in question does not exhibit arbitrage [Filipović, 2009, Ch.4.3].

Consider an asset that pays one unit at a specified future time point. Such an asset is known as a *zero-coupon bond*, and the time of its payment as its *maturity time* (note that possibility of *default* — the possibility that the payment of one unit is not made — is not accounted for here; *credit risk* is thus assumed to be absent). Let P_{tT} denote the time- t price of a zero-coupon bond maturing at T , and let $\mathbb{E}^{\mathbb{Q}}[\cdot | \mathcal{F}_t]$ denote the time- t conditional expectation operator under the risk-neutral measure (we assume here that one, denoted \mathbb{Q} , exists; the fact that several may exist is addressed below). Using the aforementioned martingale, we then have that

$$\frac{P_{tT}}{B_t} = \mathbb{E}^{\mathbb{Q}} \left[\frac{P_{TT}}{B_T} \mid \mathcal{F}_t \right],$$

which — after rearranging, using Equation 1.2, and recognising the maturity value of the zero-coupon bond ($P_{TT} = 1$) — results in

$$\begin{aligned} P_{tT} &= \mathbb{E}^{\mathbb{Q}} \left[\frac{B_t}{B_T} \mid \mathcal{F}_t \right] \\ &= \mathbb{E}^{\mathbb{Q}} [e^{-\int_t^T r_u du} \mid \mathcal{F}_t]. \end{aligned} \tag{1.3}$$

The phrase *term structure* refers to some set of related financial quantities that each correspond to a different maturity time. Although several types of quantities have a term structure (such as implied volatilities, swap rates, etc.), we use the phrase to refer specifically to zero-coupon bond prices. If one selects a particular convention to express the returns offered by investing in prevailing zero-coupon bonds until their (known) maturity payment, one gets a term structure of interest

²This of course assumes that Brownian motion pertains to the real-world (or objective) probability measure — that it could pertain to some other, risk-adjusted measure is addressed below.

³Instead of the cash account, one can use a different asset to denominate a model’s price processes, in which case the concept of a risk-neutral measure generalises to that of an *equivalent martingale measure*. The asset used as a denominator is known as the *numéraire*. While alternative choices of numéraire can be useful in certain contexts, we focus on the cash account in this thesis, and therefore on a risk-neutral measure as the particular type of equivalent martingale measure in use. A risk-neutral measure offers the helpful interpretation of being *risk-adjusted*: after averaging with these probabilities, an asset grows only as fast as the risk-free cash account (so that their ratio is a martingale), regardless of any risk premium (and therefore additional growth on average) afforded by the market.

⁴This is one aspect of the well-known *fundamental theorems of asset pricing*, outlined originally in Harrison & Kreps [1979] and Harrison & Pliska [1981], and generalised elsewhere, notably in Delbaen & Schachermayer [1994].

rates, known as the prevailing *yield curve*.

Consider the instruments central to modern interest-rate markets — coupon bonds, forward-rate agreements and swaps. Due to the simple nature of a zero-coupon bond maturity payment, these more complicated instruments are equivalent to some zero-coupon bond combination. It is therefore useful to think of zero-coupon bonds as the fundamental objects in an interest-rate market, and, from a modelling perspective, to take the term structure of zero-coupon bonds as the tradeable assets of the model. This is the set of assets considered in the definition of a risk-neutral measure and with which *portfolios* — discussed below — can be formed.

Let us suppose that the Vasiček [1977] model given by Equation 1.1 is the risk-neutral form of the model, i.e., that $\{W_t\}$ is a Brownian motion under a risk-neutral measure \mathbb{Q} (the probability measure supposed to represent objective, statistical probabilities is known as the *real-world measure* and is denoted \mathbb{P}). By virtue of satisfying Equation 1.1, the short-rate process $\{r_t\}$ exhibits the Markov property, which implies that the conditional expectation in Equation 1.3 (for given values of t and T) is given by a function of the time- t short rate — having fixed values for the model's parameters, we therefore have that

$$P_{tT} = P(t, T, r_t). \quad (1.4)$$

Notice, critically, that the existence of \mathbb{Q} has extended, via the function above, a model of a single (short-term) interest rate to one of the whole term structure. The model can be given a real-world parameterisation as well — that is, the measure appearing in Equation 1.1 can be changed (with the well-known Girsanov theorem) from \mathbb{Q} to \mathbb{P} — so that, using the above function, a (real-world) distribution for any future zero-coupon bond price or associated interest rate is implicit in the model's specification.

The risk-neutral measure is an extremely useful construct. In addition to linking zero-coupon bond prices at any particular time (with a function like that in Equation 1.4), it more generally links together the prices of all interest-rate *derivatives*: assets that, at a specified future time point, pay out an amount that depends on (is *derived* from) the term structure (at, and potentially before, the payment time). Excepting the special case of a zero-coupon bond, a derivative's pay out is not known before it is to be paid — if we let X denote an integrable random variable that represents the pay out of a particular derivative (taking place at a future time T), consider

$$\mathbb{E}^{\mathbb{Q}}[e^{-\int_t^T r_u du} X \mid \mathcal{F}_t] \quad (1.5)$$

as the derivative's time- t price. If the derivative is *attainable* — meaning that there is a self-financing portfolio of zero-coupon bonds that is guaranteed to equal the value of the derivative's pay out — the price given by Equation 1.5 is justified by the martingale relationships that prevail under \mathbb{Q} .^{5,6} Note that there may be many risk-neutral measures in a particular model, each one of which suggests a price for a certain derivative, in the manner of Equation 1.5. For an attainable derivative, these prices must converge to the cost of the *replicating* zero-coupon bond portfolio (if there are many replicating portfolios, they must have the same cost if arbitrage is to be precluded and any risk-neutral measures are to exist). If a derivative is not attainable, the range of prices across the variety of risk-neutral measures is the set of prices that are compatible with the absence

⁵It is also necessary to note that self-financing portfolios of zero-coupon bonds inherit the discounted-martingale property, as well as to apply similar manipulations to those used before Equation 1.3.

⁶Our treatment of the possible replicating portfolios — namely, that they may be comprised of zero-coupon bonds — is not universal in the literature. For example, the Vasiček [1977] model is sometimes classified as incomplete on the grounds that it, on a certain view, involves no tradeable assets with which to form replicating portfolios (see, for instance, Carmona & Tehranchi [2007, Ch.2.3]).

of arbitrage (that a derivative is not attainable does not imply that its price is unconstrained — for instance, if a derivative pay out is strictly positive, it cannot, according to any risk-neutral measure, exhibit a negative price if arbitrage is to be avoided).⁷ If all derivatives allowed in the model are attainable, the model is said to be *complete*. Term structure models driven by Brownian motion are often complete, because the martingale representation theorem can often be used to show that a replicating portfolio strategy exists for a general claim. However, as we shall see in the next subsection, this is not the case for all Brownian-based models.

In a complete model, one can calculate Equation 1.5 for a particular derivative, and interpret the result as the derivative’s theoretical, model-determined price — were its price any different, one could, according to the model’s assumptions, create an arbitrage by trading the derivative against its replicating portfolio. Moreover, this price can be calculated without explicitly determining the details of the replication strategy — the risk-neutral measure can be characterised and expressions like Equation 1.5 immediately considered.^{8,9} Note that, because of the Markov nature of most term structure models, Equation 1.5 is usually given by a function of the central stochastic variables of the model at time t — one can substitute in these known values, such as the prevailing short rate r_t , or use the function in a more abstract way, such as by translating possible future states of the world. In both cases, the relationships between the model’s prices at any particular cross section of time have been characterised. Note also that Equation 1.5 may not be amenable to an explicit calculation, and one may have to settle for some approximation.

In an incomplete model, one cannot justify a particular price for every derivative in the sense described above (namely, that the derivative must be priced equal to the cost of attaining it with other assets). A common strategy in these models is to choose one among the many risk-neutral measures, basing this choice on how the market in question has in fact priced certain derivatives (presumably within the range offered by the various risk-neutral measures). Some of the uses of the risk-neutral measure mentioned above for complete models would no longer apply, while some would. For instance, certain derivative prices would now be required as an input, and so cannot be priced with the model. There is a vast literature addressing issues arising in incomplete models, such as how replication of derivatives can be approximated in various senses, and how one can characterise and select among the various risk-neutral measures — see Björk [2004, Ch.16] (and the references therein) or, for a comprehensive treatment, Dana & Jeanblanc [2007].

The completeness or incompleteness of a term structure model is clearly one of its key characteristics, and whether one should prefer models of one type is not obvious. In the following section, we introduce a particular way in which term structure models can be made incomplete.

1.2 Unspanned Stochastic Volatility

An important aspect of a term structure model is its volatility specification: that is, the way in which the potential variation of certain quantities — particularly interest rates and bond prices — is controlled and allowed to vary. The Vasiček [1977] model involves a very simple, *constant volatility* specification — a single parameter σ controls how the variation in the short rate is derived

⁷This follows from a similar argument to that underlying Equation 1.5, but instead using portfolios that super- or sub-replicate the derivative (that is, necessarily exceed or are exceeded by the pay out).

⁸Another aspect of the fundamental theorems of asset pricing, in addition to the one mentioned in Footnote 4, is that in a complete model, not only do the pricing expressions from the various risk-neutral measures converge, the measures themselves coincide (so that the risk-neutral measure is unique).

⁹One may be interested in the nature of the underlying replicating strategy, especially if one is attempting to manage the risk associated with the derivative in question. Equation 1.5 is still useful in this regard — the sensitivity of the conditional expectation to the key factors on which it is conditioned can often imply the replicating strategy details (in simple but common settings, this is known as the *delta-hedging rule* — see Shreve [2004, Ch.4]).

from the underlying Brownian motion. A slightly more complicated volatility specification is found in the well-known [Cox, Ingersoll & Ross \[1985\]](#) model, where the short rate process $\{r_t\}$ satisfies

$$dr_t = \kappa(\theta - r_t) dt + \sigma\sqrt{r_t} dW_t,$$

where κ , θ and σ are constants and $\{W_t\}$ is a Brownian motion (to be understood as different to and separate from the one in Equation 1.1 above, giving rise to a separate short rate process; similarly, the constant parameters should be taken separately from other models). Volatility in the [Cox et al. \[1985\]](#) model is non-constant — as the short rate increases, its volatility increases with it. In this regard, one could argue that this model is more plausible than the [Vasiček \[1977\]](#) model — [Piazzesi \[2010, §7.7\]](#) documents a tendency for interest-rate volatility to be positively *level dependent* (that is, positively correlated with interest-rate levels). More complicated still is the volatility specification of the [Fong & Vasiček \[1991\]](#) model, in which the short rate $\{r_t\}$ satisfies the following coupled dynamics:

$$dr_t = \kappa_r(\theta_r - r_t) dt + \sqrt{v_t} dW_t^{(1)} \quad \text{and} \quad (1.6)$$

$$dv_t = \kappa_v(\theta_v - v_t) dt + \sigma_v\sqrt{v_t} dW_t^{(2)}, \quad (1.7)$$

where $\{W_t^{(1)}\}$ and $\{W_t^{(2)}\}$ are (possibly correlated) one-dimensional Brownian motions and where κ_r , θ_r , κ_v , θ_v and σ_v are constants. This model's volatility is not only non-constant, it is *stochastic* in its own right: there is an additional Brownian motion $\{W_t^{(2)}\}$ driving the volatility process $\{v_t\}$. Due to the autonomous stochasticity feeding into the volatility (separate to that feeding into the short rate itself), the [Fong & Vasiček \[1991\]](#) model is said to exhibit *stochastic volatility*. The volatility process $\{v_t\}$ will feature in the model's bond prices — using Equation 1.3 and the joint Markov property of $\{r_t\}$ and $\{v_t\}$, zero-coupon bonds are given by a function of the form

$$P_{tT} = P(t, T, r_t, v_t). \quad (1.8)$$

It will be useful to briefly review, and distinguish between, *local volatility* and *stochastic volatility* models in the context of single-asset models, where only one asset price process (in addition to the cash account) is modelled (in contrast to the infinitely many assets constituting the term structure). Consider firstly the well-known *local volatility model*, outlined originally by [Dupire \[1994\]](#), in which the asset price process $\{S_t\}$ satisfies¹⁰

$$dS_t = \mu S_t dt + \sigma(t, S_t) dW_t, \quad (1.9)$$

where μ is a constant, $\sigma(\cdot, \cdot)$ is a suitable real-valued function and $\{W_t\}$ is a Brownian motion. Next, consider the equally well-known [Heston \[1993\]](#) model, where the asset price process satisfies the following coupled dynamics:

$$dS_t = \mu S_t dt + \sqrt{v_t} S_t dW_t^{(1)} \quad \text{and} \\ dv_t = \kappa_v(\theta_v - v_t) dt + \sigma_v\sqrt{v_t} dW_t^{(2)},$$

where μ , κ_v , θ_v and σ_v are constants and $\{W_t^{(1)}\}$ and $\{W_t^{(2)}\}$ are (correlated) Brownian motions.

The [Fong & Vasiček \[1991\]](#) model, given by Equations 1.6 and 1.7, closely resembles the [Heston \[1993\]](#) model. There is a crucial economic difference between the two, however. *Volatility risk* in the [Fong & Vasiček \[1991\]](#) model — the uncertainty around potential changes to v_t — is *reflected*

¹⁰A discrete-time version was independently developed by [Derman & Kani \[1994\]](#).

in zero-coupon bond prices (see Equation 1.8). This risk can therefore be hedged with zero-coupon bonds, in that one can find a bond position with a value that has any desired sensitivity to the volatility process value. More generally, arbitrary sensitivities to the short rate *and* volatility processes can be acquired with suitable holdings in *two* zero-coupon bonds (that is, zero-coupon bonds of two, distinct maturities). The Fong & Vasicek [1991] model, and other term structure models exhibiting stochastic volatility, are, in this sense, unlike the Heston [1993] model, where the stochastic volatility risk cannot be (fully) hedged, because (assuming the Brownian motions are not perfectly correlated) the volatility process involves a source of variation that is separate from the underlying asset price process. In term structure models, the availability of *multiple* assets with which to form replicating portfolios causes — in the absence of certain circumstances that we address below — any stochastic volatility risk to be hedgeable with zero-coupon bonds. Because sensitivity to the volatility process can be acquired, the well-known martingale representation theorem can usually — again, modulo some special cases to be described — be used to show that a general derivative in the Fong & Vasicek [1991] model, and in term structure models based on Brownian motion, is attainable with a zero-coupon bond portfolio.

In seminal papers, Collin-Dufresne & Goldstein [2002] and Heidari & Wu [2003] present empirical evidence that some significant portion of interest-rate volatility risk cannot be hedged with zero-coupon bonds. They claim that, in order to plausibly model the interest-rate markets, term structure models should reflect this fact by exhibiting, using the terminology of Collin-Dufresne & Goldstein [2002], *unspanned stochastic volatility*; that is, a volatility specification that results in stochastic volatility risk that cannot be fully hedged with zero-coupon bonds. We now illustrate how an instance of this very specific subclass of stochastic-volatility term structure models can arise.

Thus far we have only considered models that specify dynamics for the short rate, and while this convenient in many respects, term structure models need not be constructed in this way. One alternative approach, originally due to Constantinides [1992], is to specify dynamics for the *pricing kernel process*: a strictly positive stochastic process, denoted $\{\pi_t\}$, such that the time- t price of a derivative paying out X at time T is given by

$$\frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}}[\pi_T X | \mathcal{F}_t], \quad (1.10)$$

where $\mathbb{E}^{\mathbb{P}}[\cdot | \mathcal{F}_t]$ denotes a time- t conditional expectation under the real-world measure \mathbb{P} .¹¹ Such models are known as *rational term structure models*, because Equation 1.10 shows that zero-coupon bonds (which can be considered derivatives with a deterministic pay out of one) exhibit prices that are quotients, or ratios, of expressions involving the specified pricing kernel process:

$$P_{tT} = \frac{1}{\pi_t} \mathbb{E}^{\mathbb{P}}[\pi_T P_{TT} | \mathcal{F}_t] = \frac{\mathbb{E}^{\mathbb{P}}[\pi_T | \mathcal{F}_t]}{\pi_t}. \quad (1.11)$$

Suppose that the pricing kernel process satisfies

$$d\pi_t = \kappa_{\pi}(\theta_{\pi} - \pi_t) dt + \sigma^{\pi}(\pi_t, v_t) dW_t,$$

where κ_{π} and θ_{π} are constants, $\sigma^{\pi}(\cdot, \cdot)$ is a suitable real-valued function, $\{v_t\}$ is some suitably well-behaved stochastic process and $\{W_t\}$ is a Brownian motion under the real-world measure (we assume that the volatility function $\sigma^{\pi}(\cdot, \cdot)$ keeps $\{\pi_t\}$ positive, making the above specification

¹¹Like the existence of a risk-neutral measure, the existence of a pricing kernel process (also known as the state price density) precludes arbitrage. Moreover, a complete model is synonymous with a unique pricing kernel process, in which case Equation 1.10 can be interpreted as a cost of replication.

suitable for a pricing kernel process).¹² Provided a true martingale results from the integrand process $\{\sigma^\pi(\pi_t, v_t)\}$, the well-known method of solving linear stochastic differential equations shows that Equation 1.11 simplifies to

$$P_{tT} = \frac{\theta_\pi + e^{-\kappa_\pi(T-t)}(\pi_t - \theta_\pi)}{\pi_t}.$$

Crucially, unlike in the Fong & Vasicek [1991] model, a time- t bond price is not a function of the time- t volatility process v_t . Stochastic shocks to interest-rate or bond volatility in this model do not, therefore, have an instantaneous effect on zero-coupon bond prices. Many interest-rate derivatives, such as bond options, caps and swaptions, are well known to be sensitive to volatility levels, and to thus exhibit volatility risk. This volatility risk cannot be hedged in the zero-coupon bond market according to this model, precluding certain derivatives from being attainable — the model is thus incomplete and is said to exhibit unspanned stochastic volatility. Note that recourse to multiple underlying instruments, in this or any unspanned stochastic volatility model, does not resolve the dilemma, as *all* zero-coupon bonds lack instantaneous sensitivity to the volatility process (again, this is unlike the Fong & Vasicek [1991] model, where the availability of a second bond allows the model’s volatility risk to be hedged). The volatility risk can be described as *unspanned* by the zero-coupon bonds (or by the term structure).

While stochastic volatility term structure models have some analogy to the Heston [1993] model, the analogy holds more deeply for models with unspanned stochastic volatility: there is a source of (stochastic volatility) risk that has no instantaneous effect on the fundamental asset(s) of the model (and so can only be hedged by entering the derivative market). Thus, unspanned stochastic volatility models are conceptually quite different to term structure models with (spanned) stochastic volatility, with this difference being similar to that between the Heston [1993] model and the local volatility model given by Equation 1.9 (where volatility is not constant or deterministic, but does not involve separate variation that limits the scope of the potential replicating portfolios). According to models exhibiting unspanned stochastic volatility, derivatives — whether being hedged or held naked — carry a source of risk that is distinct from the zero-coupon bond market risks. A zero-coupon bond — or any position composed of zero-coupon bonds, such as a swap — is also qualitatively different in the face of unspanned stochastic volatility, as the position’s volatility can change in a way that is unrelated to changes in its value or any changes in the yield curve.

A branch of literature studying term structure models with unspanned stochastic volatility has developed since the seminal work of Collin-Dufresne & Goldstein [2002] and Heidari & Wu [2003]. This literature can be divided, roughly but usefully, into *modelling* and *empirical* components. The former focusses on specifying, analysing and implementing models that exhibit unspanned stochastic volatility. Collin-Dufresne & Goldstein [2002] began this effort by deriving parameter restrictions on three-factor *affine term structure models* — a well-known class of short-rate models, outlined originally by Duffie & Kan [1996] and subsequently by Dai & Singleton [2000] — that result in unspanned stochastic volatility. Other modelling literature includes Casassus, Collin-Dufresne & Goldstein [2005], Jarrow, Li & Zhao [2007], Han [2007], Trolle & Schwartz [2009], Carr, Gabaix & Wu [2011], Joslin [2017], Filipović, Larsson & Statti [2017] and Filipović, Larsson & Trolle [2017].

The empirical component of the literature addresses the question of whether unspanned stochastic volatility models, in general, have empirical support. The specific question of interest is whether or not data observed in the interest-rate markets *appear to be have been realised from a model ex-*

¹²This model is reconsidered and fully specified in Example 4 in Section 3.2.

hibiting unspanned stochastic volatility; or, in other words, *whether unspanned stochastic volatility models are necessary* to plausibly model interest rates. Note that while the modelling literature is at times empirical in nature, any empirical aspects are undertaken in the context of the particular models under consideration. The question addressed by the empirical literature, on the other hand, is not concerned with specific models, and should ideally be answered in a model-independent way, or at least with only the mildest possible modelling assumptions. In addition to [Collin-Dufresne & Goldstein \[2002\]](#) and [Heidari & Wu \[2003\]](#), empirical literature includes [Fan, Gupta & Ritchken \[2003\]](#), [Li & Zhao \[2006\]](#), [Collin-Dufresne, Goldstein & Jones \[2009\]](#), [Bikbov & Chernov \[2009\]](#), [Li & Zhao \[2009\]](#) and [Andersen & Benzoni \[2010\]](#).

1.3 Thesis and Contribution Overview

This thesis is concerned with unspanned stochastic volatility (USV), in a variety of ways. Firstly, its theoretical foundation is of interest — we are particularly concerned with unifying the many extant unspanned stochastic volatility models into a common framework, and delineating features that they share. Secondly, the thesis addresses the question motivating the empirical USV literature mentioned above. Finally, we are interested in certain aspects of the modelling of USV.

The thesis proceeds as follows. The remaining subsections of this chapter summarise and discuss the content of each later chapter, and, where relevant, how their approaches and results relate to the existing literature. Chapter 2 establishes the mathematical foundation of the thesis. Based on this, Chapter 3 develops a theoretical framework in which term structure models can be characterised according to whether, and in what sense, they exhibit USV. The consequences of the several definitions that we make are explored, and the first concern mentioned above is thus addressed. Chapter 4 attends to the second concern — the necessity of USV models is empirically considered. Chapter 5 addresses the third concern; in particular, the chapter considers *bivariate* USV models.

The reader should note that Chapter 3 is the most technically and theoretically involved part of the thesis. Although Chapters 4 and 5 occasionally refer to Chapter 3, they can be read separately.

1.3.1 Chapter 2

Chapter 2 establishes the thesis’s core mathematical framework. In addition to stating and discussing the fundamental assumptions, terminology crucial to the technical aspects of the thesis is introduced. From a mathematical point of view, this core framework allows for virtually all continuous-time term structure models (TSMs) that are driven by Brownian motion, provided that they admit a *finite-dimensional state variable representation*; i.e., that zero-coupon bond (ZCB) prices are given by a function of a finite number of variables (known as state variables), such as in Equation 1.8.

The specialisation to Brownian-based processes is needed for the later chapters. Chapter 3 investigates the completeness — discussed further below — of TSMs. In order for (continuous-time) TSMs to be *potentially* complete, access to the martingale representation theorem is required. Section 3.4.3 then discusses generalising the theory of Chapter 3 to a more inclusive setting. Brownian processes are more than sufficient for the empirical concerns of Chapter 4, and in Chapter 5, the continuity of Brownian motion will prove instrumental in preventing explosions of the key processes.

It is also noteworthy that the framework, and therefore the thesis, is *classical* and *pre-crisis* in nature: by assuming that only a single term structure of default-free ZCBs prevails, we assume

the absence of credit and tenor-basis risk. While a vast literature has made explicit account for the possibility of default (see [Brigo & Mercurio \[2007, Ch.21\]](#) and the references therein), we do not address this issue. Similarly, we do not consider the growing literature that extends classical TSMs to a *post-crisis, multiple-curve setting* (see, for example, [Bianchetti & Carlicchi \[2011\]](#) or [Grbac & Runggaldier \[2015\]](#) for details).

Based on the core framework, various further definitions are developed; these are instrumental for Chapter 3 in particular. Notable amongst these is the formal notion of a complete TSM. Chapter 2 proposes two such notions. The first is orthodox: a model is considered complete if all claims — all random variables that are measurable by sigma-algebras in the model’s filtration — can be attained with a ZCB portfolio. We term this *mathematical completeness*, because of the mathematical and abstract nature of the model’s filtration, and therefore of the set of claims under consideration. The second notion of completeness only considers claims if they are measurable by sigma-algebras in the *filtration generated by ZCB price processes*. This is termed *economic completeness*, due to the necessarily economic nature of these claims — they must be a function of the yield curve and its history. This proposed distinction between the two notions of completeness features in the results of Chapter 3.

1.3.2 Chapter 3

Implicit in the USV literature — briefly reviewed in Section 1.2 — is an understanding of the USV property: in certain models, there are risk factors that have no instantaneous effect on the yield curve. These cannot be hedged with a ZCB position (or indeed any yield-curve based position, such as a coupon bond or swap portfolio), and are termed *unspanned factors*. Furthermore, these factors might affect the volatility of interest rates and bond prices, in which case they are termed *USV factors*, and the model is said to exhibit USV.

This chapter aims to develop a theoretical framework in which the USV property can be characterised formally and generally. Such a framework is not available in the existing literature. [Collin-Dufresne & Goldstein \[2002\]](#) provide a definition of *incomplete bond markets* in the context of affine term structure models, which is equivalent to the notion of unspanned factors characterised formally by [Filipović et al. \[2017\]](#), who continue to define USV factors as a certain kind of unspanned factor. The definitions of [Filipović et al. \[2017\]](#) pertain to their class of models, but they tentatively propose that their approach can be applied generally. Chapter 3 shows that their framework, despite having a wide applicability, requires generalisation. In addition to adjusting their definitions (so that they apply at least to models permitted in the framework of Chapter 2), we derive a number of ensuing results. Some of these are lemmas that exploit the properties uncovered by the definitions, and are useful tools for dealing with USV models, and also for highlighting relevant qualitative properties. Our more major results address the completeness of TSMs. We prove — modulo some technical requirements and caveats — that TSMs are mathematically complete if and only if they do not exhibit unspanned factors, and that TSMs are economically complete if and only if they do not exhibit USV. In addition to their theoretical interest, these results confirm important intuitions about unspanned factors, USV and market completeness. The presence of unspanned factors indicates an inability to fully hedge the risk sources of the model — this induces incompleteness in some sense. When these risk sources pertain to ZCB volatility, the unspanned risk is necessarily economic, and a more specific, perhaps more relevant, type of incompleteness is induced.

Our completeness results also justify the particular way in which we adjust the framework of [Filipović et al. \[2017\]](#). The approach of [Filipović et al. \[2017\]](#) is to delineate the set of directions

in which a model's state variables can move *without affecting the prices of ZCBs*. These directions form a real vector space, the dimension of which is taken as the model's number of unspanned factors. If such movements of the state variables affect ZCB volatility — in some formal sense — an unspanned factor is further classified as a USV factor. The intuition of this approach is that one can consider all of the different directions of potential state variable movement (that is, the vectors in \mathbb{R}^n , where n is the number of state variables) as the set of risks in the model. Any directions that cause no change to the term structure form a linear subspace, the dimension of which indicates the extent to which the model's risks are unspanned. Furthermore, for the models that Filipović et al. [2017] consider, a transformation can be applied to the state variables so that the unspanned directions correspond precisely to specific (transformed) state variables — then, the number of unspanned factors is the number of state variables that do not affect the term structure.

We generalise this approach in three senses. Firstly, we distinguish between *state variables* and underlying *factors*. While many models have an equal number of state variables and factors, this is not necessarily the case (see, for example, Trolle & Schwartz [2009] or Haubrich, Pennacchi & Ritchken [2012]). Allowing these numbers to differ gives our theory and results more generality and leads to an interesting category of models discussed in Section 3.4.1. Secondly, we determine the aforementioned directions *separately for each state*. We demonstrate with an example that this allows our theory to capture a type of unspanned factor that is overlooked in the framework of Filipović et al. [2017]. Thirdly, we allow for cases where the dimension of the unspanned directions varies across the state space. This leads to the possibility of a *partially spanned* factor. We introduce a criterion, based on the product of the Lebesgue and real-world probability measure, for determining whether dimensions of unspanned directions in a certain subset of the state space are significant enough to induce an unspanned factor. Critically, this criterion is found to be suitable for our theorem relating (mathematical) completeness to the absence of unspanned factors.

Using these ideas, we then formally characterise whether these unspanned factors induce stochastic volatility risk. Indeed, USV is best seen as a special case of the presence of unspanned factors — unspanned factors have no direct effect on the term structure, but may play some other role such as affecting volatility (we discuss others, including the possible redundancy of a factor). Our proposed formal notion of USV is supported by its link to (economic) completeness — we confirm that a USV factor induces incompleteness. Our converse result is only partial: some additional conditions, which hold in all existing models, are required — modulo these, we confirm that USV is the unique source of economic incompleteness in a Brownian-based TSM, and that unspanned factors themselves are not necessarily an obstacle to economic completeness. An interesting aspect of the proposed notion of USV is that it can be instantiated by models where only the correlations between ZCBs, but not the ZCB volatilities themselves, exhibit unspanned risk — this leads to the notion of *unspanned stochastic correlation*, which is new to the literature and is explored later in the chapter.

A key theme — but also the main theoretical weakness — of our approach is that certain transformations of the state variables (which create correspondence to unspanned directions, as mentioned above) are required for some of our theorems, and also for the convenient representation of the models themselves. The generalisation of our approach compared to Filipović et al. [2017] makes the existence of suitable transformations uncertain in the general case. We delineate the circumstances in which transformations are guaranteed to exist and discuss this matter in detail.

The chapter is concluded with a number of remarks and examples that ensue from the proposed theory. These include types of models that are not yet present in the literature, but are suggested by our definitions. It also includes some notes on whether, and to what extent, the scope of the theory can be increased (in particular, the inclusion of non-Brownian TSMs is addressed).

1.3.3 Chapter 4

The empirical question of whether USV models are necessary is of great academic and practical relevance. Academics are interested in whether the conceptual view of risk implied by USV models (described towards the end of Section 1.2 above) is an appropriate description of the financial markets; practitioners wish to know whether they must account for risks separate from those hedgeable in the bond and swap market. In addition, the details and challenges of model implementation hinge significantly on whether USV is incorporated. If implementing a Kalman filter, for instance, identifying any USV state variables requires that derivative price information be included in the measurements — as well as imposing data requirements, the computational burden of a likelihood maximisation is substantially increased. If USV models are unnecessary, there is no need to heed these challenges. In summary, a false positive and false negative answer to the empirical USV question would both be costly.

Chapter 4 addresses this question with a hedging approach. The general aim is to show that empirical hedging can yield a satisfactory answer. Specifically, we attempt to resolve the conflicting conclusions of Fan et al. [2003] and Li & Zhao [2006] — the two papers in the literature that have applied a hedging approach to this matter.

We implement hedges for caps written on EURIBOR and JIBAR, and also for caps simulated from two models, each with a USV and non-USV version. These simulations are inspired by Bikbov & Chernov [2009], who critique the original approaches of Collin-Dufresne & Goldstein [2002] and Heidari & Wu [2003] by determining whether their metrics can discriminate data generated from USV models versus data from standard models. Such benchmark results show whether empirical metrics have a valid implication. Our use of two, very different, benchmark model types (like our use of two datasets) increases the robustness and generality of our conclusions.

This benchmarking is our primary tool to resolve the aforementioned conflict between Fan et al. [2003] and Li & Zhao [2006]. Fan et al. [2003] find they can hedge swaptions with a high degree of effectiveness, which is the cornerstone of their case against the necessity of USV models. Li & Zhao [2006] find that while they can cross-hedge ZCBs very effectively, they cannot hedge caps or cap straddles adequately, and conclude that unspanned risks appear present in their data. Our empirical hedging effectiveness is similar to that found by Li & Zhao [2006]; however, our benchmark simulations show that the presence of USV — in the degree implied by empirically determined parameters — has a surprisingly small effect on hedging effectiveness. Observed hedging effectiveness, therefore, is not particularly informative regarding the presence of USV. Moreover, the hedging results of these two papers are compatible with results based on USV and non-USV models. We also show that bond hedging effectiveness, which was used by Li & Zhao [2006], does not make for useful comparison.

Because hedging results per se are inconclusive, we consider the *secondary tests* with which Fan et al. [2003] and Li & Zhao [2006] supplement their basic hedging results. Mostly, these are regressions that attempt to indicate the degree to which the hedge residuals can be ascribed to USV. Using our benchmark simulations, all of the previously proposed tests are shown to be unsatisfactory, and one category is shown to be severely flawed.

By scrutinising the rationale behind the best-performing secondary test (proposed by Li & Zhao [2006]), we construct an improved *two-tier* regression test, which proposes an initial regression and uses its residuals as the dependent variable in a second regression. The specific construction is designed to separate hedge errors that result from USV (if in fact present) from ones that result from an inevitably less-than-perfect hedging model specification. This construction turns out to be valid — the simulations show it to be a robust indicator of USV. When applied to the empirical

hedging errors we find, in the context of the simulation results, a strong indication of USV in both our datasets, and therefore empirical support for the USV modelling literature.

Misspecification of the hedging model, which induces hedge residuals separate from possible USV-related ones, poses an obstacle to the hedging approach. We take steps to mitigate this throughout the chapter and, in particular, account for it explicitly in the construction of the two-tier test. The chapter discusses the benefits of a hedging approach to the USV empirical question, one of which relates to the immense practical importance of hedging interest-rate derivatives: our results reflect on how the possible presence of USV manifests at the level of hedging effectiveness. We show that USV does not bear greatly on hedging effectiveness (on average — we find there to be more variation amongst data generated from USV models), but that USV is nevertheless evident in our datasets. Therefore, while the cross-sectional effects are relatively minor, USV models are necessary to plausibly model interest-rate time series.

This result, based on the two-tier test, is in alignment with the metrics of [Collin-Dufresne & Goldstein \[2002\]](#) and [Heidari & Wu \[2003\]](#), which we also implement and appraise with our benchmark simulations. They are found to be less than satisfactorily robust, but, given their relative ease of implementation, perhaps heuristically useful.

1.3.4 Chapter 5

Models with one spanned and one USV state variable — *bivariate USV models*, or, alternatively, *(1,1) models* — are the focus of Chapter 5. As the interest in USV models increases, (1,1) models become important: they are the simplest, non-trivial USV models, and are thus fundamental; they are useful for introducing and building intuitions about higher-factor models; and, when implementing higher-factor models, they often allow useful comparisons (for instance, to ascertain the contribution of additional state variables). In other words, (1,1) models — which are to USV models as one-factor models are to general models of the term structure — are important for the same reasons that one-factor models are.

The chapter begins by specialising the framework of Section 2.1 to one suitable for (1,1) models. The specialised framework allows for all bivariate models that can be cast in terms of their short rate. The special cases where USV is exhibited are then characterised, relying on the theory in Chapter 3.

Next, the central theoretical result of the chapter is presented. This theorem states, essentially, that the following conditions are equivalent for a bivariate short-rate model of the term structure:

- (A) the model has a (1,1) structure;
- (B) the bond price is an affine function of the short rate;
- (C) the short rate's risk-neutral drift is quadratic in the short rate.

With regard to (A), [Collin-Dufresne & Goldstein \[2002\]](#) purport to prove that such (1,1) models do not exist. This is mistaken — we show that counterexamples to their claim do exist, but must exhibit affine bond prices. These models, satisfying condition (B), have been considered in [Gabaix \[2009\]](#) and [Cheridito & Gabaix \[2008\]](#), where *linearity-generating* processes are defined and analysed. [Gabaix \[2009\]](#), in a time-homogeneous context, proves that (B) implies (C); [Cheridito & Gabaix \[2008\]](#) and [Carr et al. \[2011\]](#) then consider processes that conform to (C), and their generalisations. The fact that (C) implies (B) gives a way of constructing (1,1) models; that (B) implies (C) ensures that there is no alternative route of doing so. In a corollary, we show that the affine function in (B) can be attained in closed form, if a special case of (C) holds.

Interestingly, our theorem reveals that *linear-rational* models — a class of TSMs outlined by Filipović et al. [2017] — encompass a large subset of (1,1) models. Relegating calculations to an appendix, we delineate this subset by casting linear-rational models in terms of the short rate (translating from their original specification in terms of the pricing kernel). Linear-rational models have many interesting features, and we view this as one of contributions of the thesis.

We then go on to propose a specific bivariate USV model. As per our central theorem, it has affine bond prices and therefore log-affine yields, as well as a quadratic short-rate drift. We also give it a quadratic volatility, and therefore term our model the Log-Affine Double Quadratic (1,1) — or LADQ(1,1) — model. We show how the specification bounds the short rate, in a way that is necessary to avoid explosions and ensure the model is well-defined. The role of the various parameters is explained, and in particular we show how certain parameters control certain qualitative features of the yield curve and implied-volatility skew. Despite the non-standard dynamics in (C) and bond price form in (B), the model's yield curve behaviour is very similar to that in classical one-factor TSMs (such as the Vasiček [1977] and Cox et al. [1985] models). We show how a finite-difference scheme can be used to price derivatives under the LADQ(1,1) model, and also how a time inhomogeneity can be added that calibrates the model to market-prevailing yield curve.

We compare the LADQ(1,1) model to others in the literature. The most important consideration is the LRSQ(1,1) model — the specific linear-rational model outlined by Filipović et al. [2017] with a (1,1) structure. One issue is the comparison between the short-rate and pricing-kernel approaches. We show the LRSQ(1,1) model in terms of its short rate, and discuss the benefits of viewing it in this way. One is that it becomes more easily comparable to other models. Another is that (1,1) models are naturally parameterised in terms of the short rate — our main theorem shows precisely the constraints on and freedom within the (1,1) model class.

The next issue is the LADQ(1,1)-LRSQ(1,1) comparison. The LRSQ(1,1) model's first strength is its ability to exploit Fourier-based methods to price derivatives. Our finite-difference method is slower, but (in Section 5.2.4) we highlight two crucial efficiencies that can improve its implementation significantly. The second strength is the simplicity with which the model is ensured to be well-defined (the pricing kernel process must be kept positive). The LADQ(1,1) model requires an ad hoc upper bound for the short rate. We show, however, that the volatility specification that we use to do this accords with the tendency of interest-rate volatility to be positively dependent on interest-rate levels. More generally, due to the short-rate approach, the economic features of the specification (such as the parametric control mentioned above) are more transparent and amenable. Finally, a relative strength of the LADQ(1,1) model is the ability to simultaneously ensure a positive short rate and freely control the long end of the yield curve.

The chapter concludes with a brief outline of some extensions to our results.

Chapter 2

Mathematical Introduction and Foundation

Section 2.1 establishes the mathematical framework of the thesis; Section 2.2, using a number of subsections, develops further definitions that are used in Chapter 3.

2.1 Formal Framework

A *term structure model* (TSM) is characterised by two fundamental elements:

- a *state process*: some stochastic process $\{X_t\}$;
- and a *bond pricing function*: some function $P(t, T, X_t)$ with outputs that represent zero-coupon bond (ZCB) prices.

In the context of classical models without credit risk, this notion of a TSM is very general — the only notable exclusion is that of infinite-dimensional models, which we do not consider.¹

The primary goal of this section is to capture a general TSM, modulo some mathematical conditions, which are to be made explicit and discussed. We begin by formalising the two TSM elements in our first two assumptions.

Assumption 2.1.1. For some $S > 0$, $\{X_t\}_{0 \leq t \leq S}$ is an adapted, n -dimensional stochastic process, defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{0 \leq t \leq S}, \mathbb{P})$. The filtration $\{\mathcal{F}_t\}_{0 \leq t \leq S}$ is assumed to satisfy the usual conditions (of completeness and right-continuity), and to be generated by a d -dimensional (standard) Brownian motion $\{W_t^{\mathbb{P}}\}_{0 \leq t \leq S}$.

The models we consider thus have a finite time horizon S . Time-index sets will hereafter be suppressed, unless there is a specific need for clarity. The specialisation to a Brownian filtration is noteworthy — this is relied on extensively in Chapters 3 and 5. Section 3.4.3 discusses extending

¹The well-known Heath, Jarrow & Morton [1992] (HJM) models are the most important examples of infinite-dimensional TSMs. Other examples — involving an infinite-dimensional state process *and* factor process (introduced below) — are proposed and studied in Goldstein [2000] and Santa-Clara & Sornette [2001]. While these models are mathematically interesting — and the scope of Chapter 3 is limited by not considering them — they generally involve a considerable sacrifice of tractability [Ritchken & Sankarasubramanian, 1995; Chiarella & Kwon, 2003]. Implementation of numerical techniques becomes challenging and sometimes impossible, and the models become less theoretically transparent — many of the uses of term structure models discussed in Section 1.1 would not be applicable (such as the relationship between the points on the term structure discussed below Equation 1.4). This issue is revisited in Section 3.4.2, where we discuss the fact that many HJM models admit a finite-dimensional state process representation.

Chapter 3 in this vein; the introduction to Chapter 5 explains why the results of the chapter are most applicable in a Brownian framework.

The individual, one-dimensional elements of the *state process* $\{X_t\}$ are referred to as *state variables*. The i th state variable is denoted by $\{X_t^{(i)}\}$ (that is, $X_t^{(i)} = [X_t]_i$, where $[\cdot]_i$ extracts the i th entry of a row or column vector). The state process takes on values in some *state space*, denoted \mathcal{X} . Note that constraints on the state process are added in Assumption 2.1.5 below.

Assumption 2.1.2. $P : \{(t, T) | t \in [0, S], T \geq t\} \times \mathcal{X} \rightarrow \mathbb{R}^+$ is a $\mathcal{C}^{1,1,2}$ function, such that $P(T, T, x) = 1$ for all $T \in [0, S]$ and all $x \in \mathcal{X}$.

An output of the *bond pricing function* $P(t, T, X_t)$ is taken to represent the time- t price of a risk-free, unit-nominal zero-coupon bond (ZCB) that matures at time T . For any *maturity time* $T \geq 0$ and suitable *calendar time* $t \in [0, T]$, we define

$$P_{tT} = P(t, T, X_t), \quad (2.1)$$

which gives rise to a *ZCB price process* $\{P_{tT}\}_{0 \leq t \leq T}$ for each maturity time. Note that the bond pricing function is assumed to be continuously differentiable at least once in calendar and maturity time, and at least twice with respect to the state variables.

Although Assumptions 2.1.1 and 2.1.2 give the fundamental elements of a TSM, there is no guarantee that the model is well-behaved in certain, relevant senses. The following two assumptions address this by ensuring, respectively, that the short end of the term structure is amenable to financial interpretation and that the TSM precludes arbitrage opportunities. Note also that while the state process and bond pricing function are fundamental TSM elements or characteristics, they should not be viewed as TSM *ingredients* — in order to satisfy Assumption 2.1.4 below, the bond pricing function is usually *derived* (from, for example, Equation 1.3) rather than directly specified.

With the aim of characterising the short end of the term structure (that is, when calendar time and maturity time are near), define

$$r(t, x) = - \frac{\partial \log P(t, T, x)}{\partial T} \Big|_{T=t} \quad (2.2)$$

for all $t \in [0, S]$ and $x \in \mathcal{X}$. Defining $r_t = r(t, X_t)$ for all $t \in [0, S]$, we have the *short-rate process* $\{r_t\}$.² We then define

$$B_t = e^{\int_0^t r_u du}$$

for all $t \in [0, S]$, which gives rise to the *cash-account process* $\{B_t\}$.³

Assumption 2.1.3. The cash-account process $\{B_t\}$ is well-defined.

It follows that the *discounting process* $\{D_t\}$, defined with

$$D_t = \frac{1}{B_t} = e^{-\int_0^t r_u du}$$

for all $t \in [0, S]$, is well-defined. This allows us to define *discounted ZCB price processes* $\{D_t P_{tT}\}$, which are central in the following introduction of a *risk-neutral measure*.

²This is the standard no-arbitrage relation between the term structure and the short rate — see, for instance, Björk [2004, Ch.22] or Filipović et al. [2017]. This definition is in anticipation of Assumption 2.1.4; if Equation 2.2 did not hold, arbitrages between the cash account and short-term ZCBs would be possible, and the TSM in question would not satisfy Assumption 2.1.4.

³While the cash account process can be given any initial value, no generality is lost by setting $B_0 = 1$.

Assumption 2.1.4. \mathbb{Q} is a probability measure on (Ω, \mathcal{F}) , equivalent to \mathbb{P} , and discounted ZCB price processes are $(\mathbb{Q}, \{\mathcal{F}_t\})$ -martingales.

In Definitions 2.2.1 and 2.2.2 below, the notion of an *admissible portfolio*, which models a dynamic investment in the cash account and a set of ZCBs, is developed. Critically, we will see that Assumption 2.1.4 is sufficient to preclude arbitrage opportunities amongst these portfolios. A slight degree of generality is sacrificed by insisting on strict martingality of discounted ZCB price processes, which allows us to use a simple portfolio-admissibility criterion below — local martingality would similarly preclude arbitrage, but the added generality would not feature in this thesis.

Because the probability space's filtration is assumed to be generated by the Brownian motion $\{W_t^\mathbb{P}\}$, the converse of the Girsanov theorem (see Björk [2004, Ch.11]) is applicable: there exists an adapted, d -dimensional process $\{\zeta_t\}$ such that the process defined with

$$W_t = W_t^\mathbb{P} + \int_0^t \zeta_u du \quad (2.3)$$

for all $t \in [0, S]$, is a \mathbb{Q} -Brownian motion.⁴ Our final core assumption demands further structure for the state process under \mathbb{Q} .

Assumption 2.1.5. The state process $\{X_t\}$ satisfies

$$X_t = X_0 + \int_0^t \mu(u, X_u) du + \int_0^t \sigma(u, X_u) dW_u \quad (2.4)$$

for all $t \in [0, S]$, for some *drift function* $\mu : [0, S] \times \mathcal{X} \rightarrow \mathbb{R}^n$ and *diffusion function* $\sigma : [0, S] \times \mathcal{X} \rightarrow \mathbb{R}^{n \times d}$.⁵ Furthermore, $\mu(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ are differentiable.

Note that because the existence of the state process has already been assumed in Assumption 2.1.1, the usual Lipschitz and linear growth conditions are not necessary. We therefore directly impose the regularity we require (namely, differentiability of the drift and diffusion functions). Note also that, from Equation 2.4, the state process is continuous.

We refer to $\{W_t\}$ as the *factor process*, and to its individual elements $\{W_t^{(i)}\}$ as *factors*. The (n -dimensional) state process $\{X_t\}$ summarises the state of the world (by, at any given time, dictating the contemporaneous term structure via the bond pricing function, and describing the distributions of possible future states).⁶ But the fundamental stochastic driver of the state process, and therefore of the TSM, are the Brownian motions $\{W_t\}$ and $\{W_t^\mathbb{P}\}$; indeed, the state process is adapted to the filtrations generated by $\{W_t\}$ and $\{W_t^\mathbb{P}\}$ (this follows from Assumptions 2.1.5 and 2.1.1, respectively). One could argue that $\{W_t^\mathbb{P}\}$ is more fundamental and should be termed the factor process — our terminology will be more convenient, as we work primarily under the risk-neutral measure. We refer to $\{W_t^\mathbb{P}\}$ as the *real-world factor process*.

While the dimensions of the state process and the factor process are often made equal, the general TSM that satisfies the assumptions above exhibits n state variables and d factors.

From Assumptions 2.1.1 and 2.1.2, Itô's lemma is applicable to the bond pricing function. For

⁴Throughout the thesis, a Brownian motion should be understood as a standard Brownian motion, in the sense that the scalar components are mutually independent. We also note that the Radon-Nikodým derivative $\frac{d\mathbb{Q}}{d\mathbb{P}}$ is given by the Doléans exponential $\mathcal{E}\{-\int_0^t \zeta_u dW_u^\mathbb{P}\}$ at $t = S$.

⁵Note that certain authors use the term *diffusion function* to refer to $\sigma(\cdot, \cdot)$ multiplied by its transpose.

⁶Note that $\{X_t\}$ is Markov under \mathbb{Q} (by virtue of satisfying Equation 2.4) but not necessarily under \mathbb{P} (this would depend on the form of $\{\zeta_t\}$).

some $T \geq 0$, we define

$$m(t, T, x) = \frac{\partial P(t, T, x)}{\partial t} + \sum_{i=1}^n \frac{\partial P(t, T, x)}{\partial x^{(i)}} [\mu(t, x)]_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 P(t, T, x)}{\partial x^{(i)} \partial x^{(j)}} [\sigma(t, x) \sigma(t, x)^\top]_{ij}$$

for all $t \in [0, T]$ and $x \in \mathcal{X}$ (where $[\cdot]_{ij}$ extracts the entry in the i th row and j th column of a matrix). It then follows from Assumption 2.1.5 that

$$P_{tT} = P(0, T, X_0) + \int_0^t m(u, T, X_u) du + \int_0^t \nabla P(u, T, X_u)^\top \sigma(u, X_u) dW_u,$$

where $\nabla P(\cdot, \cdot, \cdot)$ denotes the gradient of the bond pricing function with respect to the state variables.⁷ Using the usual differential-form shorthand, we have

$$dP_{tT} = m(t, T, X_t) dt + \nabla P(t, T, X_t)^\top \sigma(t, X_t) dW_t. \quad (2.5)$$

We refer to this equation as the *dynamics* of $\{P_{tT}\}$, to $\{m(t, T, X_t)\}$ as its *drift process* and to $\{\nabla P(t, T, X_t)^\top \sigma(t, X_t)\}$ as its *diffusion process*. The dynamics of the corresponding discounted ZCB price process also follow from Itô's lemma:⁸

$$\begin{aligned} d(D_t P_{tT}) &= P_{tT} dD_t + D_t dP_{tT} \\ &= -P_{tT} r_t D_t dt + D_t (m(t, T, X_t) dt + \nabla P(t, T, X_t)^\top \sigma(t, X_t) dW_t) \\ &= D_t (m(t, T, X_t) - r_t P_{tT}) dt + D_t \nabla P(t, T, X_t)^\top \sigma(t, X_t) dW_t. \end{aligned}$$

Assumption 2.1.4 requires the discounted ZCB price process to be a martingale, and therefore requires that $m(t, T, X_t) = r_t P_{tT}$.⁹ The ZCB dynamics in Equation 2.5 then become

$$dP_{tT} = r_t P_{tT} dt + \nabla P(t, T, X_t)^\top \sigma(t, X_t) dW_t. \quad (2.6)$$

Using Equation 2.3, one can deduce the *real-world ZCB dynamics* (that is, ZCB dynamics based on $\{W_t^\mathbb{P}\}$ as an integrator) to be

$$dP_{tT} = (r_t P_{tT} + \nabla P(t, T, X_t)^\top \sigma(t, X_t) \zeta_t) dt + \nabla P(t, T, X_t)^\top \sigma(t, X_t) dW_t^\mathbb{P}. \quad (2.7)$$

Note that Equation 2.6 (or 2.7) imply that the ZCB price processes are semi-martingales and therefore valid integrators. Finally, compare these dynamics to that of the cash account — from its definition, we have

$$d \log(B_t) = r_t dt.$$

Itô's lemma then implies

$$dB_t = r_t B_t dt.$$

⁷That is, $\nabla P(t, T, X_t)$ is \mathbb{R}^n -valued, with

$$[\nabla P(t, T, X_t)]_i = \frac{\partial P(t, T, x)}{\partial x^{(i)}} \Big|_{x=X_t}.$$

⁸Note that $dD_t = -r_t D_t dt$. This follows from the definition of the discounting process, below Assumption 2.1.3.

⁹This must hold $\lambda \otimes \mathbb{Q}$ -almost everywhere, and therefore $\lambda \otimes \mathbb{P}$ -almost everywhere (Assumption 2.1.4). As is discussed in detail in the proofs of Chapter 3 (see Footnote 10), the two processes in question ($\{m(t, T, X_t)\}$ and $\{r_t P_{tT}\}$) can then be taken as indistinguishable.

The lack of a diffusion process and consequent lack of direct influence of the Brownian motions is noteworthy — the process lacks quadratic variation and, as its time- t gradient is \mathcal{F}_t -measurable, is locally deterministic — and is the sense in which the cash account represents a risk-free investment. At any time t , this investment offers a *local return* of r_t , as invested capital of B_t will grow by (approximately) $r_t B_t \Delta t$ over the ensuing period Δt (with this approximation converging as the period is shortened). ZCB price processes, which do exhibit a diffusion process, can be seen from Equation 2.7 to offer an *average local return* (i.e., based only on the drift process, under \mathbb{P}) that is different from the risk-free, cash-account one; in particular, $\{\zeta_t\}$ controls its excess magnitude over the risk-free return $r_t P_{tT}$ by scaling and summing the contemporaneous diffusion process.¹⁰ The process $\{\zeta_t\}$ is therefore referred to as a *market-price-of-risk* process.

2.2 Additional Definitions

Portfolios and Completeness

To construct a portfolio of ZCBs, we require a *maturity process* $\{\mathcal{T}_t\}_{0 \leq t \leq S}$ and a *holdings process* $\{\phi_t\}_{0 \leq t \leq S}$. While these processes can have any finite dimension, they will typically be d -dimensional (where d is the dimension of the factor process from Assumption 2.1.5), and so we assume this for the following exposition. The maturity process $\{\mathcal{T}_t\}$ is an \mathbb{R}^d -valued, piecewise-constant, right-continuous process that represents, at any given time t , d maturity times.¹¹ So that ZCBs with these maturity times are available for investment at each time t , we require that $[\mathcal{T}_t]_i \geq t$ for all i and $t \in [0, S]$. The holdings process $\{\phi_t\}$ represents the number of nominal units held in ZCBs with maturity times given by $\{\mathcal{T}_t\}$. Both the holdings and maturity processes are required to be adapted, so that the implied holdings can be in fact be determined at each time.

For a (d -dimensional) maturity process $\{\mathcal{T}_t\}$, we define a corresponding $d \times n$ *bond-sensitivity matrix process* $\{\nabla P_t^\mathcal{T}\}$ by specifying its entries:

$$[\nabla P_t^\mathcal{T}]_{ij} = \frac{\partial P(t, [\mathcal{T}_t]_i, x)}{\partial x^{(j)}} \Big|_{x=X_t}.$$

In addition, we let $\{P_t^\mathcal{T}\}$ be a vector process giving the time- t prices of the ZCBs with maturities given by $\{\mathcal{T}_t\}$; that is, $[P_t^\mathcal{T}]_i = P(t, [\mathcal{T}_t]_i, X_t)$. We now directly define the process that represents the value of an investment of ϕ_t units in the ZCB bonds with maturities \mathcal{T}_t for all times $t \in [0, S]$, having financed these holdings with the cash account and an initial capital of v_0 . After the definition, we motivate its particular form.

Definition 2.2.1. For some maturity process $\{\mathcal{T}_t\}$, holdings process $\{\phi_t\}$ and constant $v_0 \in \mathbb{R}$, the *portfolio value process* $\{V_t^{\mathcal{T}, \phi}\}$ is defined by

$$V_t^{\mathcal{T}, \phi} = B_t \left(v_0 + \int_0^t D_u \phi_u^\top \nabla P_u^\mathcal{T} \sigma(u, X_u) dW_u \right)$$

for all $t \in [0, S]$.

¹⁰This is expected to be a positive excess (i.e., that the ZCB drift should be greater than the cash-account drift), because of the general tendency of financial markets to be risk averse.

¹¹One could also this process to be infinite dimensional (see, for instance, Carmona & Tehranchi [2007]) and non-piecewise constant, and retain valid mathematical abstractions of typical, real-world bond portfolios (where a finite number of bonds are held for discrete periods of time). The simpler, intuitive approach will nevertheless allow replication results to be derived in Chapter 3.

To confirm that this definition is valid, firstly define

$$\psi_t = \frac{V_t^{\mathcal{T},\phi} - \phi_t^\top P_t^\mathcal{T}}{B_t}$$

for all $t \in [0, S]$, so that $V_t^{\mathcal{T},\phi} = \phi_t^\top P_t^\mathcal{T} + \psi_t B_t$, and so that ψ_t represents the number of units held in the cash account at time t . Secondly, note that

$$\begin{aligned} d(D_t V_t^{\mathcal{T},\phi}) &= d\left(v_0 + \int_0^t D_u \phi_u^\top \nabla P_u^\mathcal{T} \sigma(u, X_u) dW_u\right) \\ &= D_t \phi_t^\top \nabla P_t^\mathcal{T} \sigma(t, X_t) dW_t, \end{aligned} \quad (2.8)$$

and therefore, using a standard application of Itô's lemma, that

$$dV_t^{\mathcal{T},\phi} = r_t V_t^{\mathcal{T},\phi} dt + \phi_t^\top \nabla P_t^\mathcal{T} \sigma(t, X_t) dW_t. \quad (2.9)$$

Thirdly, using Equation 2.6 and the definitions of $\{\nabla P_t^\mathcal{T}\}$ and $\{P_t^\mathcal{T}\}$ (above Definition 2.2.1), we then have that

$$dV_t^{\mathcal{T},\phi} = r_t V_t^{\mathcal{T},\phi} dt + \phi_t^\top (dP_t^\mathcal{T} - P_t^\mathcal{T} r_t dt),$$

which, using the definition of $\{\psi_t\}$, simplifies to

$$\begin{aligned} dV_t^{\mathcal{T},\phi} &= \psi_t r_t B_t dt + \phi_t^\top dP_t^\mathcal{T} \\ &= \psi_t dB_t + \phi_t^\top dP_t^\mathcal{T}. \end{aligned}$$

This shows that the portfolio satisfies the classical *self-financing* property (see, for instance, Björk [2004, Ch.10] or Filipović [2009, Ch.4.2]). Portfolios that are not self-financing are therefore excluded by Definition 2.2.1.

Equation 2.8 shows that the discounted portfolio value process $\{D_t V_t^{\mathcal{T},\phi}\}$ is a local martingale. In the following definition, we follow Filipović [2009, Ch.4.3] by characterising the set of *admissible portfolios* based on whether $\{D_t V_t^{\mathcal{T},\phi}\}$ is a true martingale.¹²

Definition 2.2.2. A portfolio value process $\{V_t^{\mathcal{T},\phi}\}$ is *admissible* if and only if $\{D_t V_t^{\mathcal{T},\phi}\}$ is a $(\mathbb{Q}, \{\mathcal{F}_t\})$ -martingale.

An admissible portfolio value process $\{V_t^{\mathcal{T},\phi}\}$ is classically understood to be an arbitrage portfolio if and only if $V_0^{\mathcal{T},\phi} = 0$, $V_{T^*}^{\mathcal{T},\phi} \geq 0$ and $\mathbb{P}(V_{T^*}^{\mathcal{T},\phi} > 0) > 0$ for some $T^* \in [0, S]$ (see Björk [2004, Ch.7] or Filipović [2009, Ch.4.3]). If an admissible portfolio meets the first criterion, then Definition 2.2.2 implies that

$$\mathbb{E}^\mathbb{Q}[D_{T^*} V_{T^*}^{\mathcal{T},\phi}] = D_0 V_0^{\mathcal{T},\phi} = 0. \quad (2.10)$$

If it meets the second criterion, we have that $D_{T^*} V_{T^*}^{\mathcal{T},\phi} \geq 0$ (as the discounting process is positive). Together with Equation 2.10, this implies that $\mathbb{Q}(V_{T^*}^{\mathcal{T},\phi} > 0) = 0$ (because otherwise the expectation would be strictly positive). As \mathbb{Q} and \mathbb{P} are equivalent (Assumption 2.1.4), we conclude that such a portfolio cannot meet the third arbitrage criterion. TSMs with admissible arbitrage portfolios are thus excluded from our framework.

¹²This is a very direct way of excluding problematic portfolios (that is, portfolios that allow arbitrage, despite the underlying assets admitting a risk-neutral measure). There are other methods of doing so; for instance, in Björk [2004, Ch.10], the discounted value process must be bounded from below (which is a less onerous constraint than martingality).

With the notion of admissible portfolios in hand, we can now define what it means for a TSM — any state process $\{X_t\}$ and bond pricing function $P(\cdot, \cdot, \cdot)$ pair that satisfies the assumptions of Section 2.1 — to be *complete*. We define completeness in two different senses, after the following standard definitions.

Definition 2.2.3. For any $T^* \in [0, S]$, an \mathcal{F}_{T^*} -measurable random variable Θ is called a T^* -claim.

A T^* -claim represents a cash flow at time T^* ; in particular, one that may be *contingent*, in that the amount might only be revealed at the time of payment. The question of whether a claim (either particular or general) *corresponds to a contemporaneous portfolio value* is therefore of key economic interest (as discussed in Section 1.3.1).

Definition 2.2.4. A T^* -claim Θ is called *attainable* if and only if there exists an admissible portfolio value process $\{V_t^{\mathcal{T}, \phi}\}$ such that¹³

$$V_{T^*}^{\mathcal{T}, \phi} = \Theta.$$

Any admissible portfolio value process that satisfies this is said to *attain* Θ .

Definition 2.2.5. A TSM is *mathematically complete* if, for all $T^* \in [0, S]$, every T^* -claim Θ satisfying

$$\mathbb{E}^{\mathbb{P}}[|D_{T^*}\Theta|] < \infty \quad (2.11)$$

is attainable.

This is a strict notion of completeness, in that it requires attainability of all random variables that can be defined on the underlying probability space (modulo Equation 2.11). These random variables do not necessarily have economic relevance, however. A different perspective on completeness would be to consider claims only if they are measurable by the *zero-coupon bond filtration* $\{\mathcal{F}_t^P\}$: the filtration generated by all of the ZCB price processes $\{P_{tT}\}$.

Definition 2.2.6. A TSM is *economically complete* if, for all $T^* \in [0, S]$, every $\mathcal{F}_{T^*}^P$ -measurable T^* -claim satisfying Equation 2.11 is attainable.

If a claim is not based on the term structure and its history, its attainability is therefore not necessary for economic completeness.

If a TSM is mathematically complete, then \mathbb{Q} is the unique risk-neutral measure (i.e., the only measure satisfying the requirements of Assumption 2.1.4). One can see this by considering the S -claim $\frac{\mathbb{I}_A}{D_S}$, where A is some generic element of \mathcal{F}_S (and where $\mathbb{I}_{\{\cdot\}}$ denotes the indicator function), which clearly satisfies Equation 2.11. Assuming mathematical completeness, there exists an admissible portfolio value process $\{V_t^{\mathcal{T}, \phi}\}$ attaining this claim, and we have that

$$V_0^{\mathcal{T}, \phi} = D_0 V_0^{\mathcal{T}, \phi} = \mathbb{E}^{\mathbb{Q}}[D_S V_S^{\mathcal{T}, \phi}] = \mathbb{E}^{\mathbb{Q}}[\mathbb{I}_A] = \mathbb{Q}(A).$$

If $\bar{\mathbb{Q}}$ were also a risk-neutral measure, admissible portfolio value processes would also be $(\bar{\mathbb{Q}}, \{\mathcal{F}_t\})$ -martingales.¹⁴ We would then have that

$$\mathbb{Q}(A) = D_0 V_0^{\mathcal{T}, \phi} = \mathbb{E}^{\bar{\mathbb{Q}}}[D_S V_S^{\mathcal{T}, \phi}] = \bar{\mathbb{Q}}(A),$$

¹³As is standard, equality of random variables is meant in the sense of \mathbb{P} -almost sure equality.

¹⁴Portfolio value processes would certainly be $(\bar{\mathbb{Q}}, \{\mathcal{F}_t\})$ -local martingales, and they would be $(\bar{\mathbb{Q}}, \{\mathcal{F}_t\})$ -martingales if and only if they were $(\mathbb{Q}, \{\mathcal{F}_t\})$ -martingales (i.e., if they were admissible).

which shows that \mathbb{Q} and $\bar{\mathbb{Q}}$ agree on any event of the probability space. If a TSM is economically complete, one can only conclude, using the same reasoning, that all risk-neutral measures agree on the probability of events contained in \mathcal{F}_S^P .

Transformations

Two types of transformations are used in Chapter 3; the first is a *state process transformation*.

Definition 2.2.7. For an invertible function $A : \mathcal{X} \rightarrow \hat{\mathcal{X}}$, the *transformed state process* $\{\hat{X}_t\}$ is defined with $\hat{X}_t = A(X_t)$ for all $t \in [0, S]$. The codomain $\hat{\mathcal{X}}$ is termed the *transformed state space*. Letting $B : \hat{\mathcal{X}} \rightarrow \mathcal{X}$ denote the inverse transformation (so that $B(A(x)) = x$ for all $x \in \mathcal{X}$), the *transformed bond pricing function* is defined and denoted with

$$\hat{P}(t, T, \hat{x}) = P(t, T, B(\hat{x}))$$

for all $T \geq 0$, $t \in [0, T]$ and $\hat{x} \in \hat{\mathcal{X}}$.

Crucially, a TSM is *observationally equivalent* after a state process transformation — using the transformed state process and transformed bond pricing function, we have

$$\begin{aligned} \hat{P}(t, T, \hat{X}_t) &= P(t, T, B(\hat{X}_t)) \\ &= P(t, T, B(A(X_t))) \\ &= P(t, T, X_t) \\ &= P_{tT}. \end{aligned}$$

The sensitivities of the transformed bond pricing function can be related, with the chain rule, to pre-transformation sensitivities — assuming that $B(\cdot)$ is differentiable at $\hat{x} \in \hat{\mathcal{X}}$, we have

$$\begin{aligned} \frac{\partial \hat{P}(t, T, \hat{x})}{\partial \hat{x}^{(i)}} &= \frac{\partial P(t, T, B(\hat{x}))}{\partial \hat{x}^{(i)}} = \sum_{k=1}^n \frac{\partial P(t, T, x)}{\partial x^{(k)}} \bigg|_{x=B(\hat{x})} \frac{\partial [B(\hat{x})]_k}{\partial \hat{x}^{(i)}} \\ &= \nabla P(t, T, B(\hat{x}))^\top \frac{\partial B(\hat{x})}{\partial \hat{x}^{(i)}}. \end{aligned} \quad (2.12)$$

It is convenient, for a particular transformation, to define the *Jacobian matrix*; define an $n \times n$ matrix-valued function $J_A(\cdot)$ with entries given by

$$[J_A(x)]_{ij} = \frac{\partial [A(x)]_i}{\partial x^{(j)}}$$

for all $x \in \mathcal{X}$ for which $A(\cdot)$ is differentiable. The Jacobian of the inverse function, denoted $J_B(\cdot)$, is defined in the analogous way.

Assuming, for some $x \in \mathcal{X}$, that $A(\cdot)$ is differentiable at x and $B(\cdot)$ is differentiable at $A(x)$, we have that

$$J_A(x)^{-1} = J_B(A(x)),$$

as a straightforward consequence of the chain rule.¹⁵ Similarly,

$$J_A(B(\hat{x}))^{-1} = J_B(\hat{x}) \quad (2.13)$$

¹⁵One can verify this by differentiating either side of the identity $x = B(A(x))$. More generally, if $A(\cdot)$ and $B(\cdot)$ were not necessarily inverses, one would have $J_{B \circ A}(x) = J_B(A(x))J_A(x)$.

for any $\hat{x} \in \hat{\mathcal{X}}$ that allows the necessary differentiability.

Suppose that the k th component of $A(\cdot)$, which we denote $A^{(k)}(\cdot)$, is \mathcal{C}^2 . Then Itô's lemma can be used to deduce the dynamics of the k th transformed state variable $\{\hat{X}_t^{(k)}\}$ as

$$d\hat{X}_t^{(k)} = \nabla A^{(k)}(X_t)^\top dX_t + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 A^{(k)}(x)}{\partial x^{(i)} \partial x^{(j)}} \Big|_{x=X_t} [\sigma(t, X_t) \sigma(t, X_t)^\top]_{ij} dt. \quad (2.14)$$

In addition to these state process transformations, we require transformations of the factor process.

Definition 2.2.8. Suppose that $\epsilon : \mathcal{X} \rightarrow \mathbb{R}^{d \times d}$ is such that $\epsilon(x)$ is an orthogonal matrix for all $x \in \mathcal{X}$ (that is, the columns of $\epsilon(x)$ form an orthonormal basis for \mathbb{R}^d). Then the *transformed factor process*, denoted $\{\hat{W}_t\}$, is defined by¹⁶

$$\hat{W}_t = \int_0^t \epsilon(X_u)^{-1} dW_u \quad (2.15)$$

for all $t \in [0, S]$. Similarly, the *transformed real-world factor process*, denoted $\{\hat{W}_t^\mathbb{P}\}$, is defined by

$$\hat{W}_t^\mathbb{P} = \int_0^t \epsilon(X_u)^{-1} dW_u^\mathbb{P} \quad (2.16)$$

for all $t \in [0, S]$.

Note that this *orthonormal factor process transformation* is different in nature to the state process transformation in Definition 2.2.7; the latter assigns a new value to each possible state process value, while the former constructs a new value at each time point by integrating over the history of the factor process. Each infinitesimal change in the factor process implies a certain infinitesimal change in the transformed factor process; the factor process transformation is thus *local* in nature.

Importantly, using Lévy's characterisation (see Karatzas & Shreve [2012, Ch.3.3]), one can conclude that $\{\hat{W}_t\}$ is a \mathbb{Q} -Brownian motion. First consider its quadratic variation:

$$\begin{aligned} d[\hat{W}, \hat{W}]_t &= \epsilon(X_t)^{-1} (\epsilon(X_t)^{-1})^\top dt \\ &= \epsilon(X_t)^{-1} \epsilon(X_t) dt \\ &= I_d dt, \end{aligned}$$

where I_d denotes the $d \times d$ identity matrix (note that the first step above follows from the central property of orthogonal matrices; namely, that $\epsilon(x)^{-1} = \epsilon(x)^\top$). In addition to this, it suffices to note that $\{\hat{W}_t\}$ is a zero-initialised, continuous $(\mathbb{Q}, \{\mathcal{F}_t\})$ -local martingale. Using the same reasoning, $\{\hat{W}_t^\mathbb{P}\}$ is a \mathbb{P} -Brownian motion.

Pushforward Measures

Letting λ denote the Lebesgue measure, we give the final definition of the chapter, which will be used extensively in Chapter 3.

¹⁶Note that, because $\epsilon(x)^{-1} = \epsilon(x)^\top$, the orthogonality of $\epsilon(x)$ ensures the square-integrability of $\{\epsilon(X_t)^{-1}\}$. Also, we will consider measurable functions $\epsilon(\cdot)$, ensuring that $\{\epsilon(X_t)^{-1}\}$ is measurable and therefore, up to a modification, progressive [Karatzas & Shreve, 2012, Ch.1.1].

Definition 2.2.9. Let $\lambda \otimes \mathbb{P}$ denote the product of λ and \mathbb{P} , and let $X_*(\lambda \otimes \mathbb{P})$ denote the pushforward of $\lambda \otimes \mathbb{P}$ over $X(\cdot, \cdot)$ (where $X(t, \omega) = X_t$) to \mathcal{X} ; that is, for all $E \subseteq \mathcal{X}$,

$$\begin{aligned} X_*(\lambda \otimes \mathbb{P})(E) &= \lambda \otimes \mathbb{P}(X^{-1}(E)) \\ &= \lambda \otimes \mathbb{P}\left(\{(t, \omega) \in [0, S] \times \Omega \mid X_t \in E\}\right). \end{aligned}$$

Intuitively, a $X_*(\lambda \otimes \mathbb{P})$ -non-null set has some probability of containing the state process for non-negligible (λ -non-null) period of time.

Finally, suppose that a state process transformation, as per Definition 2.2.7, has been applied. Then a separate pushforward measure $\hat{X}_*(\lambda \otimes \mathbb{P})$ — which measures subsets of the transformed state space $\hat{\mathcal{X}}$ — is easily defined by analogy to Definition 2.2.9.

Chapter 3

Unspanned Factor Theory

In addition to assuming the framework of Section 2.1, which formally introduced a TSM as a combination of a bond pricing function and state process, this chapter will rely on the following two assumptions. The first specialises Assumptions 2.1.2 and 2.1.5 to be *time homogeneous*.

Assumption 3.0.1. $P : \mathbb{R}^+ \times \mathcal{X} \rightarrow \mathbb{R}^+$ is a $\mathcal{C}^{1,2}$ function such that

$$P_{tT} = P(T - t, X_t) \quad (3.1)$$

for all $t \in [0, S]$ and $T \geq t$. Furthermore, the state process $\{X_t\}$ satisfies

$$X_t = X_0 + \int_0^t \mu(X_u) du + \int_0^t \sigma(X_u) dW_u \quad (3.2)$$

for all $t \in [0, S]$, where $\mu : \mathcal{X} \rightarrow \mathbb{R}^n$ and $\sigma : \mathcal{X} \rightarrow \mathbb{R}^{n \times d}$ are differentiable drift and diffusion functions, respectively.

Time homogeneity thus entails ZCB prices that do not depend on calendar time t and maturity time T separately, and drift and diffusion functions without an explicit dependence on time. Time-inhomogeneous models would complicate the chapter substantially, and are addressed in Section 3.4.2.

We then assume the following regarding the pushforward measure from Definition 2.2.9.

Assumption 3.0.2. The Lebesgue measure λ , restricted to \mathcal{X} , is absolutely continuous with respect to the pushforward measure $X_*(\lambda \otimes \mathbb{P})$; that is, if $X_*(\lambda \otimes \mathbb{P})(E) = 0$ for some $E \subseteq \mathcal{X}$, then $\lambda(E) = 0$.

Provided that the state space \mathcal{X} is not unnecessarily large, the above assumption holds for the Brownian-based state processes used in financial modelling: any $E \subseteq \mathcal{X}$ that is λ -non-null (again, provided it is not in an extraneous part of \mathcal{X}) will have some probability of containing the state process.

The chapter proceeds as follows. Section 3.1 develops a theoretical framework to rigorously characterise unspanned factors, after which Section 3.2 considers the roles that such factors can play (including, notably, a stochastic-volatility role). Section 3.3 contains the proofs of the completeness results in the earlier sections. Section 3.4 gives some remarks and examples that follow from the chapter.

3.1 Unspanned Factors

ZCB prices depend — we have assumed — on the state process. While a change in the state process will typically cause a change in the term structure, certain state process movements might cause no such change. We begin by defining, for each state, the set of directions in which the state process can move *without affecting the term structure*.

Definition 3.1.1. For any $x \in \mathcal{X}$, the *unspanned state directions* are given by

$$\mathcal{S}^x = \{\delta \in \mathbb{R}^n \mid \nabla P(\tau, x)^\top \delta = 0 \quad \forall \tau > 0\}.$$

For some δ to be included in \mathcal{S}^x , the δ -directional derivative of the state- x bond pricing function (namely, $\nabla P(\tau, x)^\top \delta$) must be zero *for all maturities* $\tau > 0$. More abstractly, \mathcal{S}^x is the intersection, over all $\tau > 0$, of the kernels of each $1 \times n$ matrix $\nabla P(\tau, x)^\top$.¹ Movements of the state process in such a direction, by this definition, have no local effect on the yield curve. We refer to these directions as *unspanned*.

Example 1. Consider the [Fong & Vasicek \[1991\]](#) model given by Equations 1.6 and 1.7, where $\{W_t^{(1)}\}$ and $\{W_t^{(2)}\}$ are the two components of a \mathbb{Q} -Brownian motion $\{W_t\}$, and where the processes satisfying Equations 1.6 and 1.7 constitute the two-dimensional state process. The bond pricing function is of the form

$$P(\tau, x) = \exp \left(a(\tau) - b_r(\tau)x^{(1)} - b_v(\tau)x^{(2)} \right),$$

where $a(\cdot)$, $b_r(\cdot)$ and $b_v(\cdot)$ are deterministic functions derived in [Selby & Strickland \[1995\]](#) (the [Fong & Vasicek \[1991\]](#) model is an affine term structure model; these models feature bond pricing functions that are *exponential-affine* in their state variables). Although $b_v(\cdot)$ is a complicated function, it is easy to verify that it is *not* proportional to $b_r(\cdot)$; that is, one cannot find a constant β such that

$$b_v(\tau) = \beta b_r(\tau)$$

for all $\tau > 0$. Because of the form of the bond pricing function, this implies that the entries of $\nabla P(\tau, x)$, for any $x \in \mathcal{X}$, are similarly not proportional, which in turn implies that the model has no non-trivial unspanned state directions (i.e., $\mathcal{S}^x = \{\mathbf{0}_2\}$ for all $x \in \mathcal{X}$). To see this, note that for a particular maturity $\tau^* > 0$ (and a particular state x), there does exist $\delta^* \in \mathbb{R}^n$ such that²

$$\nabla P(\tau^*, x)^\top \delta^* = 0,$$

but, because $\nabla P(\tau, x)$ is not proportional across maturities, the above equation would not hold for all $\tau > 0$, and δ^* would not be an element of \mathcal{S}^x . Intuitively, one can note that arbitrary sensitivities to both state variables $x^{(1)}$ and $x^{(2)}$ *can* be acquired with two ZCBs, *because* the entries of $\nabla P(\tau, x)$ are not proportional. If the entries of $\nabla P(\tau, x)$ were proportional across maturities, any sensitivity one acquired to $x^{(1)}$ would force one's hand in acquiring $x^{(2)}$ -sensitivity (regardless of the maturities considered). This would indicate that certain movements in the state process do not affect the

¹Any particular function of an n -dimensional vector, such as $P(\tau^*, \cdot)$ for some $\tau^* > 0$, will have $(n-1)$ dimensions of unspanned directions in the sense of Definition 3.1.1; i.e., there will be $(n-1)$ linearly independent directions in which the function does not change (these directions are given by the kernel of $\nabla P(\tau^*, x)^\top$). The critical aspect of the definition is the intersection over all maturities $\tau > 0$ — as further maturities are checked, many of the $(n-1)$ dimensions will fail to remain in the intersection.

²For example, one could set $\delta^* = [b_v(\tau^*) \quad -b_r(\tau^*)]^\top$.

term structure (and therefore such movements cannot be hedged with ZCBs). The Fong & Vasicek [1991] model, however, is free of such unspanned state directions.

The state variables are, in a certain sense, less fundamental to a TSM than the factors that underlie them. Roughly speaking, movements in the factor process $\{W_t\}$ are transmitted to the state process via the volatility function $\sigma(\cdot)$ in Equation 3.2. In the same spirit as Definition 3.1.1, the following definition captures directions in which the factor process can move without any local effect on the term structure.

Definition 3.1.2. For any $x \in \mathcal{X}$, the *unspanned factor directions* are given by

$$\mathcal{U}^x = \{\delta \in \mathbb{R}^d \mid \nabla P(\tau, x)^\top \sigma(x) \delta = 0 \quad \forall \tau > 0\}.$$

There are two ways for a non-zero $\delta \in \mathbb{R}^d$ to be an element of \mathcal{U}^x . Firstly, it can *correspond to a non-trivial unspanned state direction*, provided that direction lies in the image of $\sigma(x)$ — if $\delta^* \in \text{Im}(\sigma(x))$, then there exists $\delta \in \mathbb{R}^d$ such that $\delta^* = \sigma(x)\delta$; if δ^* is also an element of \mathcal{S}^x , it follows from the definition that $\delta \in \mathcal{U}^x$.

Secondly, $\delta \in \mathbb{R}^d$ might lie in the kernel of $\sigma(x)$; that is, $\sigma(x)\delta = \mathbf{0}_n$, which ensures that $\delta \in \mathcal{U}^x$. The vector δ then corresponds to the trivial unspanned state direction $\mathbf{0}_n \in \mathcal{S}^x$.

Whether non-trivial unspanned factor directions exist, and whether many different (that is, linearly independent) such directions exist simultaneously, will prove an important characteristic of a TSM. It is therefore useful to partition the state space according to the *dimension* of the unspanned direction set (which is easily seen to be a real vector space of dimension no greater than d).

Definition 3.1.3. The \mathcal{U} -*partition* of the state space is given by

$$\mathcal{X}_{\mathcal{U}}^{(i)} = \{x \in \mathcal{X} \mid \dim(\mathcal{U}^x) = i\}$$

for $i = 0, 1, \dots, d$.

TSMs very often exhibit a diffusion function $\sigma(\cdot)$ that outputs invertible matrices $\sigma(x)$ (this involves an equal number of factors and state variables, so that $n = d$ and $\sigma(x)$ is square, in addition to $\sigma(x)$ needing to be full rank). In these cases, the unspanned factor and state direction sets \mathcal{U}^x and \mathcal{S}^x are closely linked, in that each element in one corresponds to an element in the other. This is because the kernel of $\sigma(x)$ is trivial in these cases (ruling out the second type of unspanned factor direction described above), and the image of $\sigma(x)$ covers all of \mathbb{R}^n (causing all unspanned state directions to induce a corresponding unspanned factor direction). This causes \mathcal{U}^x and \mathcal{S}^x to have an equal dimension for all $x \in \mathcal{X}$ and there is no need to consider them separately. In the general case, it is necessary to distinguish between \mathcal{U}^x and \mathcal{S}^x . In particular, the dimension of \mathcal{U}^x across the state space — captured in Definition 3.1.3 — will prove more fundamental to the TSM, especially to its potential completeness. In Section 3.4.1, we will show that the second type of unspanned factor directions (those lying in the kernel of $\sigma(x)$) — which are usually avoided by the specification of an invertible $\sigma(x)$ — can play an interesting role in a TSM.

Consider now an orthonormal factor process transformation, as described in Definition 2.2.8. The following result shows how a suitable transformation can make the unspanned factor directions *correspond to particular (transformed) factors*.

Lemma 3.1.4. There exists an orthonormal factor process transformation such that ZCB dynamics can be given in terms of the transformed factor process $\{\hat{W}_t\}$ and a modified volatility function

$\hat{\sigma}(\cdot)$:

$$dP_{tT} = r_t P_{tT} dt + \nabla P(T-t, X_t)^\top \hat{\sigma}(X_t) d\hat{W}_t,$$

where

$$[\nabla P(\tau, x)^\top \hat{\sigma}(x)]_i = 0 \quad \forall \tau > 0, \forall x \in \bigcup_{j=d-i+1}^d \mathcal{X}_{\mathcal{U}}^{(j)}$$

for $i = 1, 2, \dots, d$. Real-world ZCB dynamics, in terms of $\{\hat{W}_t^{\mathbb{P}}\}$, are then given by

$$dP_{tT} = (r_t P_{tT} + \nabla P(T-t, X_t)^\top \sigma(X_t) \zeta_t) dt + \nabla P(T-t, X_t)^\top \hat{\sigma}(X_t) d\hat{W}_t^{\mathbb{P}}.$$

Remark 3.1.5. Lemma 3.1.4 reveals the implications of the unspanned factor direction sets \mathcal{U}^x being non-trivial. When $X_t \in \mathcal{X}_{\mathcal{U}}^{(i)}$ (i.e., when the prevailing unspanned factor direction set is i -dimensional), the final i transformed factors are not needed as integrators for the next infinitesimal change in ZCB prices (because the corresponding entries of the ZCB diffusion process are equal to zero). In other words, when $X_t \in \mathcal{X}_{\mathcal{U}}^{(i)}$, the final i elements of the stochastic shock $d\hat{W}_t$ do not transmit through to the term structure.

Proof. For each $x \in \mathcal{X}$, define a $d \times d$ matrix $\epsilon(x)$ such that the right-most columns of $\epsilon(x)$ form an orthonormal basis for \mathcal{U}^x , and the remaining columns form an orthonormal basis for $(\mathcal{U}^x)^\perp$ (where $(\mathcal{U}^x)^\perp$ denotes the orthogonal complement of \mathcal{U}^x with respect to \mathbb{R}^d). For all $x \in \mathcal{X}$, therefore, the columns of $\epsilon(x)$ form an orthonormal basis for \mathbb{R}^d , and $\epsilon(\cdot)$ is suitable for inducing an orthonormal factor process transformation as outlined in Definition 2.2.8. Defining the transformed factor process $\{\hat{W}_t\}$ as in Equation 2.15, consider the ZCB dynamics from Equation 2.6 (with Assumption 3.0.1 applied):

$$\begin{aligned} dP_{tT} &= r_t P_{tT} dt + \nabla P(T-t, X_t)^\top \sigma(X_t) dW_t \\ &= r_t P_{tT} dt + \nabla P(T-t, X_t)^\top \sigma(X_t) \epsilon(X_t) \epsilon(X_t)^{-1} dW_t \\ &= r_t P_{tT} dt + \nabla P(T-t, X_t)^\top \sigma(X_t) \epsilon(X_t) d\hat{W}_t. \end{aligned} \tag{3.3}$$

Defining $\hat{\sigma}(X_t) = \sigma(X_t) \epsilon(X_t)$, we get the dynamics written in the lemma. Because the last columns of $\epsilon(x)$ are elements of \mathcal{U}^x , the last elements of

$$\nabla P(T-t, x)^\top \hat{\sigma}(x) = \nabla P(T-t, x)^\top \sigma(x) \epsilon(x)$$

are equal to zero (this follows directly from Definition 3.1.2). Specifically, the last element is equal to zero wherever $\dim(\mathcal{U}^x) > 0$; that is, when $x \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}$. Similarly, the second last element is equal to zero wherever $\dim(\mathcal{U}^x) > 1$; that is, when $x \in \bigcup_{j=2}^d \mathcal{X}_{\mathcal{U}}^{(j)}$. Continuing, one gets the form written in the lemma. The real-world ZCB dynamics follow from updating the dynamics in Equation 2.7 (under Assumption 3.0.1) with Equation 2.16, exactly as done above for the risk-neutral dynamics. \square

We now assume that the above basis vectors can be chosen such that the resultant $\{\epsilon(X_t)\}$ is suitably regular. Such a choice is easy to identify in practice, as is shown and discussed in Example 3 below.

Assumption 3.1.6. One can choose $\epsilon : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $\{\epsilon(X_t)\}$ is $\lambda \otimes \mathbb{P}$ -almost everywhere left-continuous.

Having applied Lemma 3.1.4, the dynamics of the original state process can be updated — introducing the transformed factor process to Equation 3.2, we get

$$dX_t = \mu(X_t) dt + \hat{\sigma}(X_t) d\hat{W}_t. \quad (3.4)$$

Following on from Remark 3.1.5, suppose, for instance, that $\mathcal{X}_{\mathcal{U}}^{(1)}$ occupies the whole state space. Then ZCBs can be written as a stochastic integral against the first $d - 1$ transformed factors, and ZCB prices are never sensitive to changes in the final transformed factor. Such a model can therefore be said to exhibit an unspanned factor. The following definitions characterise the general case, where the dimension of \mathcal{U}^x may vary over the state space.

Definition 3.1.7. Let p equal the largest value of i for which $\mathcal{X}_{\mathcal{U}}^{(i)}$ is $X_*(\lambda \otimes \mathbb{P})$ -non-null. The TSM in question is said to have p *unspanned factors*.

Definition 3.1.8. Let q equal the smallest value of i for which $\mathcal{X}_{\mathcal{U}}^{(i)}$ is $X_*(\lambda \otimes \mathbb{P})$ -non-null. The TSM in question is said to have q *fully unspanned factors*, and $p - q$ *partially spanned factors* (where p is determined in Definition 3.1.7).

Suppose, for instance, that $\mathcal{X}_{\mathcal{U}}^{(0)}$ and $\mathcal{X}_{\mathcal{U}}^{(1)}$ partition the state space into $X_*(\lambda \otimes \mathbb{P})$ -non-null blocks. While such a model has one unspanned factor by Definition 3.1.7, we can see that this unspanned factor (i.e., the single dimension of unspanned directions) prevails *in only part of the state space*. We term this a *partially spanned factor* — the second transformed factor will be required as an integrator only in part of the state space (when $X_t \in \mathcal{X}_{\mathcal{U}}^{(0)}$). We now consider TSMs that exhibit unspanned factors.

Example 2. Consider a TSM where the (risk-neutral) dynamics of the state process are given by

$$dX_t = \begin{bmatrix} -\kappa X_t^{(1)} \\ -\kappa X_t^{(2)} \end{bmatrix} dt + \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} dW_t \quad (3.5)$$

for non-zero constants κ , σ_1 and σ_2 (and for a two-dimensional Brownian motion $\{W_t\}$). If the short rate is given by $r_t = \varphi + X_t^{(1)} + X_t^{(2)}$, for some $\varphi \in \mathbb{R}$, the resulting bond pricing function is³

$$P(\tau, x) = \exp \left(a(\tau) - \frac{1 - e^{-\kappa\tau}}{\kappa} x^{(1)} - \frac{1 - e^{-\kappa\tau}}{\kappa} x^{(2)} \right),$$

where $a(\cdot)$ is a deterministic function. This model is a special case of the G2++ model outlined by Brigo & Mercurio [2007, Ch.4]; in particular, we have taken a time-homogeneous version and have forced the two state variables to have equal rates of mean reversion. The G2++ model also allows for a correlation parameter, but this does not affect the relevant aspects of the example and has been set to zero.

Like in Example 1, the bond-pricing function is exponential-affine. Here, however, there is proportionality (indeed, equality) between the affine coefficient functions — because of the equal mean-reversion rates, the partial derivatives of the bond pricing function at *any* maturity are equal for either state variable. One therefore finds non-trivial unspanned state directions; for all $x \in \mathcal{X} = \mathbb{R}^2$, we have

$$\mathcal{S}^x = \left\{ \begin{bmatrix} \alpha & -\alpha \end{bmatrix}^\top \mid \alpha \in \mathbb{R} \right\}.$$

³That is, $P(\cdot, \cdot)$ returns this short rate function (via the bond-short rate relationship in Equation 2.2) and satisfies Equation 1.3 (and thus gives rise to discounted ZCB price processes that are \mathbb{Q} -martingales).

State process movements in these directions *preserve the sum of the two state variables*, and therefore have no effect on ZCB prices (as the bond pricing function above depends only on the sum of the state variables).

The square and full-rank diffusion matrix in Equation 3.5 causes the unspanned factor direction set \mathcal{U}^x to be of equal dimension to \mathcal{S}^x . The vectors in \mathcal{U}^x differ slightly to those in \mathcal{S}^x (as transmission through the diffusion matrix needs to be accounted for) — we have

$$\mathcal{U}^x = \left\{ \begin{bmatrix} \alpha\sigma_2 & -\alpha\sigma_1 \end{bmatrix}^\top \mid \alpha \in \mathbb{R} \right\}$$

for all $x \in \mathcal{X}$. \mathcal{U}^x is always one-dimensional, so we have $\mathcal{X}_\mathcal{U}^{(1)} = \mathcal{X}$; applying Definition 3.1.7, this TSM contains one unspanned factor, because the only higher-dimensional block of the \mathcal{U} -partition, $\mathcal{X}_\mathcal{U}^{(2)}$, is empty and therefore $X_*(\lambda \otimes \mathbb{P})$ -null. From Definition 3.1.8 (because $\mathcal{X}_\mathcal{U}^{(0)}$ is empty), the model exhibits one fully unspanned factor, or, in other words, the TSM's one unspanned factor is fully unspanned.

In order to apply Lemma 3.1.4, it is useful to define

$$\bar{\sigma}_i = \frac{\sigma_i}{\sqrt{(\sigma_1)^2 + (\sigma_2)^2}} \quad (3.6)$$

for $i = 1$ and 2 . Then consider a constant function given by

$$\epsilon(x) = \begin{bmatrix} \bar{\sigma}_1 & \bar{\sigma}_2 \\ \bar{\sigma}_2 & -\bar{\sigma}_1 \end{bmatrix}$$

for all $x \in \mathcal{X}$. This function is suitable for the orthonormal transformation — the right-hand column is a unit-length element of \mathcal{U}^x for all x , and the left-hand column is orthogonal (making it a basis for $(\mathcal{U}^x)^\perp$) and also of unit length. The process $\{\epsilon(X_t)\}$ is used to define a transformed factor process $\{\hat{W}_t\}$ as per the proof of Lemma 3.1.4. Following from Equation 3.3, ZCB dynamics are given by

$$\begin{aligned} dP_{tT} &= r_t P_{tT} dt + \nabla P(T-t, X_t)^\top \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \bar{\sigma}_1 & \bar{\sigma}_2 \\ \bar{\sigma}_2 & -\bar{\sigma}_1 \end{bmatrix} d\hat{W}_t \\ &= r_t P_{tT} dt + \nabla P(T-t, X_t)^\top \begin{bmatrix} \sigma_1 \bar{\sigma}_1 & \sigma_1 \bar{\sigma}_2 \\ \sigma_2 \bar{\sigma}_2 & -\sigma_2 \bar{\sigma}_1 \end{bmatrix} d\hat{W}_t, \end{aligned}$$

which, using Equation 3.6 and the bond pricing function above, simplifies to

$$dP_{tT} = r_t P_{tT} dt + \left[\frac{P_{tT}}{\kappa} (e^{-\kappa\tau} - 1) \sqrt{(\sigma_1)^2 + (\sigma_2)^2} \quad 0 \right] d\hat{W}_t.$$

Lemma 3.1.4 makes a TSM's unspanned factors explicit — the unspanned factor directions are isolated and consigned to the final transformed factors. In addition to *illustrating* the effect of unspanned factors, this can be *useful to implement* — having applied Lemma 3.1.4, ZCB dynamics in Example 2 are simpler and more transparent, and many further applications of the orthonormally transformed Brownian motions follow below.

Example 2 interacts with Definitions 3.1.7 and 3.1.8 in a straightforward way, in that $\mathcal{X}_\mathcal{U}^{(1)}$ occupies all of the state space, while $\mathcal{X}_\mathcal{U}^{(0)}$ and $\mathcal{X}_\mathcal{U}^{(2)}$ are empty. Note that if $\mathcal{X}_\mathcal{U}^{(0)}$ and $\mathcal{X}_\mathcal{U}^{(2)}$ were non-empty but $X_*(\lambda \otimes \mathbb{P})$ -null, the definition would return the same conclusion (that one fully spanned factor is exhibited). This criterion of $X_*(\lambda \otimes \mathbb{P})$ -non-nullness, used in Definitions 3.1.7 and 3.1.8, is demonstrated in the following example.

Example 3. In a particular *quadratic term structure model* (see Leippold & Wu [2002] and Ahn, Dittmar & Gallant [2002]), the two-dimensional state process satisfies

$$dX_t = \begin{bmatrix} -\kappa X_t^{(1)} \\ -\kappa X_t^{(2)} \end{bmatrix} dt + \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix} dW_t, \quad (3.7)$$

where κ and σ are non-zero constants, and the short rate is modelled with

$$r_t = X_t^\top \begin{bmatrix} \psi & 0 \\ 0 & \psi \end{bmatrix} X_t + \varphi$$

for constants φ and $\psi \neq 0$. The TSM's bond pricing function is given by

$$P(\tau, x) = \exp \left(a(\tau) - b(\tau)(x^{(1)})^2 - b(\tau)(x^{(2)})^2 \right),$$

where $a(\cdot)$ and $b(\cdot)$ are deterministic functions satisfying ordinary differential equations (ODEs) given in Leippold & Wu [2002].⁴

For all $x \in \mathcal{X} \setminus \{\mathbf{0}_2\}$, it is easy to verify that

$$\mathcal{S}^x = \left\{ \begin{bmatrix} \alpha x^{(2)} & -\alpha x^{(1)} \end{bmatrix}^\top \mid \alpha \in \mathbb{R} \right\}.$$

If the *sum of the squared state variables* has a particular value, the bond pricing function returns the same prices regardless of the individual states — if state process values are plotted on a Cartesian plane, one can imagine concentric circles that preserve the sum of the squared state variables. The vectors in \mathcal{S}^x are the local directions of movement that keep any state x on a particular circle.

For the $X_*(\lambda \otimes \mathbb{P})$ -null state $x = \mathbf{0}_2$, which is a stationary point of the bond pricing function at all maturities, we have $\mathcal{S}^x = \mathbb{R}^2$.

The TSM's diffusion matrix (which is a scaled identity matrix) results in the unspanned factor directions \mathcal{U}^x being precisely equal to the unspanned state directions \mathcal{S}^x for all $x \in \mathcal{X}$, so that $\mathcal{X}_\mathcal{U}^{(1)} = \mathcal{X} \setminus \{\mathbf{0}_2\}$ and $\mathcal{X}_\mathcal{U}^{(2)} = \{\mathbf{0}_2\}$. From Definition 3.1.7, the TSM exhibits one unspanned factor — although the largest value of i for which $\mathcal{X}_\mathcal{U}^{(i)}$ is *non-empty* is 2, in order to get $X_*(\lambda \otimes \mathbb{P})$ -non-nullness, one needs to set $i = 1$. Because $i = 1$ cannot be decreased any further without getting $X_*(\lambda \otimes \mathbb{P})$ -nullness ($\mathcal{X}_\mathcal{U}^{(0)}$ is empty and therefore null), Definition 3.1.8 stipulates that the TSM contains one fully unspanned factor.

The use of $X_*(\lambda \otimes \mathbb{P})$ -non-nullness in Definitions 3.1.7 and 3.1.8 is natural in the context of the completeness results later in this and the next section. It is also intuitive; for an unspanned factor to be declared, an unspanned factor dimension needs, with non-zero probability, to prevail for some appreciable period — not merely some instant — of time. In this example, a second unspanned factor dimension arises when $X_t = \mathbf{0}_2$; but, because there is no probability of the state occupying this value for any *period* (for any λ -non-null set of time points), it is appropriate for Definitions 3.1.7 to reflect only one unspanned factor.

⁴The absence of certain terms in the bond pricing function is essential for the example. In a general quadratic term structure model with two state variables, the bond pricing function is given by

$$P(\tau, x) = \exp \left(a(\tau) - x^\top \begin{bmatrix} b_1(\tau) & b_2(\tau) \\ b_3(\tau) & b_4(\tau) \end{bmatrix} x - x^\top \begin{bmatrix} c_1(\tau) \\ c_2(\tau) \end{bmatrix} \right).$$

Our particular specification of the short rate results in $b_2(0) = b_3(0) = c_1(0) = c_2(0) = 0$; our particular specification of the mean-reversion matrix ensures that these functions do not change from their initial conditions as τ increases from zero. The symmetry of our specification results in $b_1(\tau) = b_4(\tau)$ for all $\tau > 0$, although a more general specification could give $b_1(\tau) = \alpha b_4(\tau)$, for some $\alpha \neq 0$, which would preserve all the relevant aspects of the example.

Consider defining

$$\epsilon(x) = \begin{bmatrix} \frac{x^{(1)}}{\sqrt{(x^{(1)})^2 + (x^{(2)})^2}} & \frac{x^{(2)}}{\sqrt{(x^{(1)})^2 + (x^{(2)})^2}} \\ \frac{x^{(2)}}{\sqrt{(x^{(1)})^2 + (x^{(2)})^2}} & -\frac{x^{(1)}}{\sqrt{(x^{(1)})^2 + (x^{(2)})^2}} \end{bmatrix} \quad (3.8)$$

for all $x \in \mathcal{X} \setminus \{\mathbf{0}_2\}$. For all of these states, the right-hand column is a unit-length element of \mathcal{U}^x , while the left-hand column is unit-length and orthogonal. We can also set $\epsilon(\mathbf{0}_2)$ as the 2×2 identity matrix I_2 (or as any other orthonormal basis of \mathbb{R}^2). Then $\{\epsilon(X_t)\}$ is suitable for Lemma 3.1.4 — ZCB dynamics, calculated as per Equation 3.3, would only require the first transformed factor $\{\hat{W}_t^{(1)}\}$ as an integrator.

Note also that Assumption 3.1.6 is satisfied — $\{\epsilon(X_t)\}$ is $\lambda \otimes \mathbb{P}$ -almost everywhere left-continuous (as the function given in Equation 3.8, which covers $X_*(\lambda \otimes \mathbb{P})$ -almost all of the state space, is continuous). In general, one would expect the continuity of the state process to allow an orthonormal basis to be chosen in a sufficiently regular way.

In a certain sense, unspanned factors appear pathological, especially fully unspanned factors — one might think that a d -factor TSM with p unspanned factors should simply be recast as a $(d-p)$ -factor model with no unspanned factors. However, a TSM cannot necessarily be recast equivalently in this way; the unspanned factors do not influence the ZCB prices directly, but might influence the model in some other way, for example by playing a role in the volatility specification. We now point out that unspanned factors — whether or not they play any type of role — cause a TSM to be mathematically incomplete, and, with some assumption of hedging maturity availability, are the unique such cause.

Theorem 3.1.9. Given Assumptions 3.1.6 and 3.3.2, a TSM is mathematically complete if and only if it has no unspanned factors.

The proof is given in Section 3.3.1, as is Assumption 3.3.2. While the absence of unspanned factors ensures that suitable ZCB maturities for a general replicating portfolio can be selected at all times $t \in [0, S]$, Assumption 3.3.2 ensures that maturities can be selected in advance, for intervals of time (as is required by our ZCB portfolio construction in Definition 2.2.1). This assumption holds in all cases of practical interest, and is discussed in Section 3.3.1 and Remark 3.3.9.

While unspanned factors induce mathematical incompleteness, we shall see that they do not necessarily induce economic incompleteness — this will hinge on the particular effect the unspanned factors have on the model in question, which is the concern of Section 3.2.

The connection of Theorem 3.1.9 to the second fundamental theorem of asset pricing [Delbaen & Schachermayer, 1994] is elegant and intuitive — a TSM will admit a unique risk-neutral measure if and only if it lacks unspanned factors. Unspanned factors, if present, give the freedom for multiple risk-neutral measures to exist.

Remark 3.1.10. Filipović et al. [2017] characterise unspanned factors by defining (modulo some notational differences) the following set:

$$\{\delta \in \mathbb{R}^n \mid \nabla P(\tau, x)^\top \delta = 0 \quad \forall \tau > 0, \forall x \in \mathcal{X}\}.$$

This set's dimension is taken as the number of unspanned factors. As discussed in Section 1.3.2, our approach generalises this in three senses. Firstly, we allow for separate treatment of the state variables and underlying factors (which, as discussed below Definition 3.1.3, is necessary in the event of a diffusion function $\sigma(\cdot)$ that is not square and full rank). Secondly, we *do not require*

particular unspanned directions to prevail in multiple states; unlike Filipović et al. [2017], our unspanned factor directions are defined individually for each state. This allows unspanned factor and state directions to be *state dependent* — if, for instance, \mathcal{U}^x is one-dimensional for all $x \in \mathcal{X}$ (so that the TSM exhibits one unspanned factor and, according to Theorem 3.1.9, is mathematically incomplete), it makes no difference in our theory if \mathcal{U}^x varies with x (like it does in Example 3).⁵ Finally, we do not require dimensions of unspanned directions to prevail in all states. This gives rise to our notion of a partially spanned factor, and allows us to effectively ignore $X_*(\lambda \otimes \mathbb{P})$ -null states in Definitions 3.1.7 and 3.1.8. Theorem 3.1.9 provides rationale to our proposed definitions — the use of the pushforward measure from Definition 2.2.9 turns out to link directly to completeness or lack thereof.

Consider now the state variables. An analogue to the partition in Definition 3.1.3 is easily defined.

Definition 3.1.11. The \mathcal{S} -partition of the state space is given by

$$\mathcal{X}_S^{(i)} = \{x \in \mathcal{X} \mid \dim(\mathcal{S}^x) = i\}$$

for $i = 0, 1, \dots, n$.

If $\sigma(x)$ is not invertible for all $x \in \mathcal{X}$, the \mathcal{S} -partition will not coincide with the \mathcal{U} -partition. We therefore summarise this partition separately, but in precisely the same vein as the unspanned factor directions.

Definition 3.1.12. Let \bar{p} equal the largest value of i for which $\mathcal{X}_S^{(i)}$ is $X_*(\lambda \otimes \mathbb{P})$ -non-null. The TSM in question is said to have \bar{p} *unspanned state variables*.

Definition 3.1.13. Let \bar{q} equal the smallest value of i for which $\mathcal{X}_S^{(i)}$ is $X_*(\lambda \otimes \mathbb{P})$ -non-null. The TSM in question is said to have \bar{q} *fully unspanned state variables*, and $\bar{p} - \bar{q}$ *partially spanned state variables* (where \bar{p} is determined in Definition 3.1.12).

Like the unspanned factor directions in Definitions 3.1.7 and 3.1.8, the unspanned state variables referred to in the definitions are *latent*: a transformation is required for the *dimensions of unspanned state directions to correspond to specific state variables*. While Lemma 3.1.4 achieves this for factors, its state-process analogue is below termed \mathcal{S} -transformation; this is a particular type of state process transformation described in Definition 2.2.7. Unlike the orthonormal factor transformation in Lemma 3.1.4, however, an \mathcal{S} -transformation is not guaranteed to exist. A requirement of Definition 2.2.7 is that a state process transformation is invertible — this is essential to maintaining the state-based description of the TSM in terms of the transformed state process $\{\hat{X}_t\}$. This is an onerous requirement in general, but appears to be met in all cases in the existing literature. This issue is discussed in Remark 3.1.15 below.

We now define the class of transformations that make the unspanned state directions correspond to specific transformed state variables. The definition includes some technical requirements, which are discussed below.

Definition 3.1.14. Suppose a state process transformation $A(\cdot)$ and its inverse $B(\cdot)$ are $X_*(\lambda \otimes \mathbb{P})$ - and $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere differentiable, that the first $n - \bar{q}$ components of $A(\cdot)$ are \mathcal{C}^2 , and that the transformed state space $\hat{\mathcal{X}}$ is convex. Suppose furthermore that, for $i = n - \bar{q} + 1, \dots, n$,

⁵See Remark 3.3.3 for a more specific discussion of why the capturing of state dependent directions is necessary for Theorem 3.1.9, or indeed for any general link between unspanned factors and incompleteness.

the transformed bond pricing function $\hat{P}(\cdot, \cdot) = P(\cdot, B(\cdot))$ satisfies

$$\frac{\partial \hat{P}(\tau, \hat{x})}{\partial \hat{x}^{(i)}} = 0 \quad \forall \tau > 0, \quad (3.9)$$

for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all $\hat{x} \in \hat{\mathcal{X}}$. Then the state process transformation is called an \mathcal{S} -transformation.

The primary requirement of Definition 3.1.14 is the lack of sensitivity of the transformed bond pricing function specified by Equation 3.9, which is applied to the final \bar{q} transformed state variables (\bar{q} is the TSM's number of fully unspanned state variables, from Definition 3.1.13). Intuitively, the transformation must make the \bar{q} dimensions of unspanned state directions (that prevail $X_*(\lambda \otimes \mathbb{P})$ -almost everywhere) correspond to the final \bar{q} transformed state variables.

Almost everywhere differentiability of the transformation is a mild requirement; the stronger \mathcal{C}^2 requirement on the earlier components of $A(\cdot)$ ensure that Itô's lemma can be applied to the state variables that do not correspond to fully unspanned state directions. The requirement of convexity is met in practice (and is discussed further in the proof of Lemma 3.1.17).

Example 2 (Continued). The unspanned state directions in Example 2 are given by

$$\mathcal{S}^x = \left\{ \begin{bmatrix} \alpha & -\alpha \end{bmatrix}^\top \mid \alpha \in \mathbb{R} \right\} \quad (3.10)$$

for all $x \in \mathcal{X}$. This is everywhere one-dimensional, so that $\mathcal{X}_S^{(1)} = \mathcal{X}$; applying Definitions 3.1.12 and 3.1.13, the TSM exhibits one (fully) unspanned state variable. As expected (due to the square and full-rank diffusion matrix of the model), this is equal to the number of (fully) unspanned factors.

Now consider the following (invertible) state process transformation:

$$\hat{X}_t = A(X_t) = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} X_t \Rightarrow B(\hat{X}_t) = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \hat{X}_t. \quad (3.11)$$

The inverse gives the transformed bond pricing function as

$$\begin{aligned} \hat{P}(\tau, \hat{x}) &= P(\tau, B(\hat{x})) = \exp \left(V(\varphi, \tau) - \frac{1 - e^{-\kappa\tau}}{\kappa} \hat{x}^{(2)} - \frac{1 - e^{-\kappa\tau}}{\kappa} (\hat{x}^{(1)} - \hat{x}^{(2)}) \right) \\ &= \exp \left(V(\varphi, \tau) - \frac{1 - e^{-\kappa\tau}}{\kappa} \hat{x}^{(1)} \right). \end{aligned} \quad (3.12)$$

One can then see that

$$\frac{\partial \hat{P}(\tau, \hat{x})}{\partial \hat{x}^{(2)}} = 0 \quad \forall \tau > 0,$$

for all $\hat{x} \in \hat{\mathcal{X}}$, satisfying the primary requirement of Definition 3.1.14. The one unspanned state direction now appears explicitly, as ZCBs are insensitive to one of the transformed state variables. In addition, the transformation poses no differentiability problems, and results in a convex state space $\hat{\mathcal{X}} = \mathbb{R}^2$, confirming that the above $A(\cdot)$ induces an \mathcal{S} -transformation.

To see why the above state process transformation was suitable, consider

$$\nabla \hat{P}(\tau, \hat{x})^\top = \nabla P(\tau, B(\hat{x}))^\top J_B(\hat{x}), \quad (3.13)$$

which follows from Equation 2.12 when applied for all values of i simultaneously. Equation 3.9 in

Definition 3.1.14 requires that the final \bar{q} entries of the left-hand side are equal to zero ($\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere). This is achieved if and only if the final \bar{q} columns of $J_B(\hat{x})$ are elements of $\mathcal{S}^{B(\hat{x})}$ (the unspanned state directions at $x = B(\hat{x})$, given by Definition 3.1.1), as $\mathcal{S}^{B(\hat{x})}$ is precisely the kernel of $\nabla P(\tau, B(\hat{x}))^\top$ for all maturities $\tau > 0$. In Example 2 above, we have

$$J_B(\hat{x}) = \begin{bmatrix} 0 & 1 \\ 1 & -1 \end{bmatrix} \quad (3.14)$$

for all $\hat{x} \in \hat{\mathcal{X}}$. The second column is an element of $\mathcal{S}^{B(\hat{x})}$ (see Equation 3.10), ensuring that Equation 3.9 is satisfied. Definition 3.1.14 places no requirements on the first column — it must simply maintain invertibility of the matrix and therefore of the whole transformation.

Remark 3.1.15. It is highly preferable for a TSM to admit an \mathcal{S} -transformation. The ability to link any unspanned factors — which, as mentioned above, are not necessarily pathological or redundant — to specific state variables (not merely to general state process directions) is instrumental from a theoretical perspective (as will become evident in this and the following section).⁶ It can be practically useful as well — certain qualitative aspects of a model become much easier to discern when the unspanned directions are explicitly reflected. For example, the transformed bond pricing function can give valuable intuition on how the term structure is in fact susceptible to change (as this can be obscured if the unspanned directions are not consigned to specific state variables).

The crucial variable in identifying an \mathcal{S} -transformation is whether the unspanned state directions are constant across the state space (that is, whether $\mathcal{S}^{x_1} = \mathcal{S}^{x_2}$ for $x_1, x_2 \in \mathcal{X}$). In these cases, a *linear* \mathcal{S} -transformation — of the form $\hat{X}_t = A(X_t) = AX_t$, for $A \in \mathbb{R}^{n \times n}$ — exists. As illustrated by Example 2, the coefficient matrix of the inverse transformation can be set by setting the \bar{q} right-most columns as any basis for \mathcal{S}^x , and the remaining columns as any vectors that result in a full-rank matrix (see Equations 3.10 and 3.14). The inverse of this matrix can then be used to define a linear \mathcal{S} -transformation. Notably, the invertibility and differentiability constraints are not problematic in this setting.

However, as mentioned in Remark 3.1.10, our definitions allow for \mathcal{S}^x to vary with x . In these cases, a non-linear \mathcal{S} -transformation is necessary, as the right-most columns of $J_B(\hat{x})$ cannot be constant (if they are to lie in the state-dependent \mathcal{S}^x). While these cases necessitate more complicated \mathcal{S} -transformations and are less convenient to work with, they do nevertheless give rise to unspanned state variables and factors and therefore a mathematically incomplete TSM — they must be captured by our definitions in order to prove Theorem 3.1.9 (and the theorems in Section 3.2). We do not address the existence of an \mathcal{S} -transformation in a general case in this chapter, and do so briefly in Appendix A, where we show why a suitable and intuitive \mathcal{S} -transformation should exist *in principle*, at least in a local sense. While global and closed-form \mathcal{S} -transformations are not guaranteed, a practical strategy for identifying a transformation is given in Appendix A — this generalises the method used for Example 3 below.

All TSMs in the existing USV literature have constant unspanned state directions, and thus admit a linear \mathcal{S} -transformation. One should certainly prefer with model classes that admit convenient \mathcal{S} -transformations — an interesting open question is whether models with non-constant unspanned state directions could be desirable. Perhaps, in certain circumstances, such models could exhibit benefits that trump the difficulties of the necessarily non-linear transformations.

⁶This is not to say that a TSM should necessarily be cast initially in an \mathcal{S} -transformed form. There can be technical reasons to prefer a specification of the state process that requires a transformation; for example, Filipović et al. [2017] specify a state process that, although it requires a transformation, admits a regularly shaped state space that helps to keep the model well-defined.

Example 3 (Continued). Consider the following state process transformation:

$$A^{(1)}(x) = (x^{(1)})^2 + (x^{(2)})^2, \quad (3.15)$$

and

$$A^{(2)}(x) = \begin{cases} 0 & \text{if } x = \mathbf{0}_2, \\ \text{atan2}(x^{(2)}, x^{(1)}) & \text{otherwise.} \end{cases} \quad (3.16)$$

This is the polar-coordinate transformation, except that $A^{(1)}(x)$ represents the *squared* distance of x from the origin (rather than the distance itself), ensuring that $A^{(1)}(\cdot)$ is \mathcal{C}^2 (note also that $A^{(2)}(\cdot) = 0$ is differentiable everywhere except on the $X_*(\lambda \otimes \mathbb{P})$ -null subset $\{[\alpha \ 0]^\top \mid \alpha \leq 0\}$). An (almost everywhere differentiable) inverse is admitted, namely

$$B(\hat{x}) = \begin{bmatrix} \sqrt{\hat{x}^{(1)}} \cos(\hat{x}^{(2)}) \\ \sqrt{\hat{x}^{(1)}} \sin(\hat{x}^{(2)}) \end{bmatrix}$$

for all $\hat{x} \in \mathbb{R}^+ \times (-\pi, \pi] = \hat{\mathcal{X}}$, which is convex. The transformed bond pricing function is then given by

$$\begin{aligned} \hat{P}(\tau, \hat{x}) &= P(\tau, B(\hat{x})) = \exp \left(a(\tau) - b(\tau) \hat{x}^{(1)} (\cos^2(\hat{x}^{(2)}) + \sin^2(\hat{x}^{(2)})) \right) \\ &= \exp \left(a(\tau) - b(\tau) \hat{x}^{(1)} \right), \end{aligned} \quad (3.17)$$

which can be seen to satisfy Equation 3.9 in Definition 3.1.14, as the transformed bond pricing function has a partial derivative of zero with respect to the final transformed state variable.

Remark 3.1.16. Although our definitions allow for partially spanned factors and state variables (Definitions 3.1.8 and 3.1.13, respectively), models exhibiting them do not, to the author's knowledge, appear in the literature. We will therefore largely assume their absence until Section 3.4.1, where they are specifically discussed. This is sometimes for notational convenience — the sacrificed generality is minor, and additional notation could easily be added if necessary. Some of the results below, though, do in fact depend on the absence of partially spanned factors and state variables, but because existing models conform to this absence, the applicability of these results is not greatly vitiated.

If an \mathcal{S} -transformation is possible and there are no partially spanned state variables, it is useful to partition the \mathcal{S} -transformed state process into separate components. Define two sub-processes of the transformed state process as follows:

$$\hat{X}_t = A(X_t) = \begin{bmatrix} Y_t \\ U_t \end{bmatrix}, \quad (3.18)$$

where $\{Y_t\}$ is the $(n - \bar{p})$ -dimensional, *spanned component*, and $\{U_t\}$ is the \bar{p} -dimensional, *fully unspanned component* of the transformed state process (recall that \bar{p} is the number of unspanned state variables from Definition 3.1.12; this is now assumed to also be the number of fully spanned state variables). In these cases, it is also useful to partition the transformed state space $\hat{\mathcal{X}}$ into components corresponding to $\{Y_t\}$ and $\{U_t\}$; that is, $\hat{\mathcal{X}} = \mathcal{Y} \times \mathcal{W}$, where $\mathcal{Y} \subseteq \mathbb{R}^{(n - \bar{p})}$ and $\mathcal{W} \subseteq \mathbb{R}^{\bar{p}}$.⁷

⁷This assumes that $\hat{\mathcal{X}}$ is a product of $(n - \bar{p})$ - and \bar{p} -dimensional spaces. This assumption is not necessary for our

In both Examples 2 and 3, ZCB prices can be given by the spanned component only (see Equations 3.12 and 3.17). This is very useful and holds more generally — to see this, define a function $\tilde{P} : \mathbb{R}^+ \times \mathcal{Y} \rightarrow \mathbb{R}^+$ with

$$\tilde{P}(\tau, y) = \hat{P}\left(\tau, \begin{bmatrix} y \\ \bar{u} \end{bmatrix}\right) \quad (3.19)$$

for all $\tau > 0$ and $y \in \mathcal{Y}$, where \bar{u} is some element of \mathcal{W} . We term $\tilde{P}(\cdot)$ the *reduced bond pricing function*, because the following lemma confirms that it gives bond prices in terms of only the spanned component $\{Y_t\}$.

Lemma 3.1.17. Assuming that there are no partially spanned state variables and that an \mathcal{S} -transformation is possible, we have

$$P_{tT} = \tilde{P}(T - t, Y_t)$$

for all $t \in [0, S]$ and $T \geq t$.

Proof. Let \bar{u} be some fixed element of \mathcal{W} . For general $y \in \mathcal{Y}$ and $u \in \mathcal{W}$, define $\gamma^{y,u} : [0, 1] \rightarrow \mathbb{R}^n$ with

$$\gamma^{y,u}(s) = \begin{bmatrix} y \\ u \end{bmatrix} s + \begin{bmatrix} y \\ \bar{u} \end{bmatrix} (1 - s)$$

for all $s \in [0, 1]$. This gives a bijective parameterisation of the straight curve with endpoints $[y \ \bar{u}]^\top = \gamma^{y,u}(0)$ and $[y \ u]^\top = \gamma^{y,u}(1)$. Note that due to the convexity condition in Definition 3.1.14, this curve does not leave $\hat{\mathcal{X}}$, so that $\hat{P}(\tau, \gamma^{y,u}(s))$ is well-defined.⁸

Using this curve and the fundamental theorem of calculus, for all $\tau > 0$, we have that

$$\begin{aligned} \hat{P}\left(\tau, \begin{bmatrix} y \\ u \end{bmatrix}\right) - \hat{P}\left(\tau, \begin{bmatrix} y \\ \bar{u} \end{bmatrix}\right) &= \hat{P}(\tau, \gamma^{y,u}(1)) - \hat{P}(\tau, \gamma^{y,u}(0)) \\ &= \int_0^1 \frac{\partial \hat{P}(\tau, \gamma^{y,u}(s))}{\partial s} ds \\ &= \int_0^1 \nabla \hat{P}(\tau, \gamma^{y,u}(s))^\top \frac{d\gamma^{y,u}(s)}{ds} ds \\ &= \int_0^1 \nabla \hat{P}(\tau, \gamma^{y,u}(s))^\top \begin{bmatrix} \mathbf{0}_{(n-\bar{p})} \\ u - \bar{u} \end{bmatrix} ds. \end{aligned} \quad (3.20)$$

From Definition 3.1.14, the above integrand is $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere equal to zero (without partially spanned factors, Definition 3.1.14 requires that the final \bar{p} elements of $\nabla \hat{P}(\tau, \hat{x})$ are $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere equal to zero). By Assumption 3.0.2, the integrand is therefore λ -almost everywhere equal to zero, causing the integral to have a value of zero.⁹ The lemma follows

results, but it assists our notation greatly and is met in all cases in practice (Appendix A shows why the transformed state space should indeed be a product, and also be convex — see Footnote 8).

⁸Even though Appendix A shows that the convexity requirement of Definition 3.1.14 would be met, it could be relaxed if necessary — a path-connectedness requirement on \mathcal{W} would suffice.

⁹This is an intermediate step required for Assumption 3.0.2 to be directly applicable. Letting $\hat{E} \subseteq \hat{\mathcal{X}}$ denote the transformed state process values where the integrand is non-zero, it follows immediately that $B(\hat{E})$ is $X_*(\lambda \otimes \mathbb{P})$ -null. We then have

$$\lambda(\hat{E}) = \int_{\hat{E}} \lambda(d\hat{x}) = \int_{B(\hat{E})} \|J_A(x)\| \lambda(dx) = 0,$$

where the final step follows directly from Assumption 3.0.2.

from considering the left-hand side and comparing to Equation 3.19. \square

Without any partially spanned state variables, the number of unspanned state variables \bar{p} is equal to the number of fully unspanned state variables \bar{q} . Because an \mathcal{S} -transformation requires the first $n - \bar{p}$ components of $A(\cdot)$ to be \mathcal{C}^2 , Itô's lemma is applicable to the spanned component $\{Y_t\}$. Using Equation 2.14 (which gives the dynamics of a particular transformed state variable in terms of dX_t) and Equation 3.4 (which gives dX_t in terms of the transformed factor process $d\hat{W}_t$), one can deduce dynamics for $\{Y_t\}$. Moreover, one can define a drift function $\mu^Y(\cdot)$ and a diffusion function $\hat{\sigma}(\cdot)$ such that

$$dY_t = \mu^Y(\hat{X}_t) dt + \hat{\sigma}(\hat{X}_t) d\hat{W}_t, \quad (3.21)$$

because it follows from Equations 2.14 and 3.4 that the drift and diffusion processes are only functions of X_t , and then the inverse function $B(\cdot)$ can be used to define $\mu^Y(\cdot)$ and $\hat{\sigma}(\cdot)$. In the case of the latter, we have

$$\hat{\sigma}(\hat{x}) = J_{\bar{A}}(B(\hat{x}))\hat{\sigma}(B(\hat{x})), \quad (3.22)$$

where $\bar{A}(\cdot)$ denotes the first $n - \bar{p}$ components of $A(\cdot)$, so that $Y_t = \bar{A}(X_t)$ (and that $J_{\bar{A}}(\cdot)$ returns an $(n - \bar{p}) \times n$ matrix of partial derivatives). The following lemma shows that this diffusion function must be of a certain form, namely that its final p columns are equal to zero (recall that p is the number of unspanned factors from Definition 3.1.7).

Lemma 3.1.18. Assuming there are no partially spanned state variables or factors and that an \mathcal{S} -transformation is possible, the spanned component $\{Y_t\}$ of the transformed state process then satisfies

$$dY_t = \mu^Y(\hat{X}_t) dt + \begin{bmatrix} \sigma^Y(\hat{X}_t) & \mathbf{0}_{(n-\bar{p}) \times p} \end{bmatrix} d\hat{W}_t,$$

where $\sigma^Y(\hat{x})$ is defined as the first $d - p$ columns of $\hat{\sigma}(\hat{x})$. Furthermore, $\sigma^Y(\hat{x})$ has at least as many rows as it has columns (i.e., $n - \bar{p} \geq d - p$), and is $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere full rank.

Remark 3.1.19. Recall that the orthonormal factor process transformation of Lemma 3.1.4 isolated the unspanned factor directions by making them correspond to the final transformed factors, the stochastic shocks of which do not enter ZCB price processes. Lemma 3.1.18 shows that this is reflected intuitively at the level of the state process: these shocks do not enter the spanned component (if they are to have any effect, it can only be via the unspanned component). The result is useful as it allows $\{Y_t\}$ to be written without involving the final p elements of $\{\hat{W}_t\}$ as integrators.

The size constraint on $\sigma^Y(\hat{x})$ is also intuitive: an excess of columns would imply that there are more spanned factors than spanned state variables — this is untenable because it would force $\sigma^Y(\hat{x})$ to exhibit a non-trivial kernel, which would show that some of the putative spanned factors are in fact unspanned. More generally, an excess of factors over state variables results in part of the factor process being unspanned and redundant.

Proof. To justify the primary claim, we need to prove that the final p columns of $\hat{\sigma}(\hat{x})$ are equal to zero for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all \hat{x} .¹⁰ To do this, we will show that the rows of $J_{\bar{A}}(B(\hat{x}))$ in Equation

¹⁰This implies that the final p columns of $\{\hat{\sigma}(\hat{X}_t)\}$ are $\lambda \otimes \mathbb{P}$ -almost everywhere zero. They can therefore be taken as indistinguishable from zero — any $\lambda \otimes \mathbb{P}$ -null changes to stochastic integrands do not cause any changes to the

3.22 are orthogonal to the final p columns of $\hat{\sigma}(B(\hat{x}))$ for all \hat{x} , excepting some $\hat{X}_*(\lambda \otimes \mathbb{P})$ -null subsets of $\hat{\mathcal{X}}$ that are excluded during the proof. Three claims are necessary to establish this.

Firstly, any of the $n - \bar{p}$ rows of $J_{\bar{A}}(B(\hat{x}))$ are orthogonal to any of the final \bar{p} columns of $J_B(\hat{x})$ (and therefore to the linear span of these columns) for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all \hat{x} . This follows from Equation 2.13: we have that $J_A(B(\hat{x}))J_B(\hat{x}) = I_n$ wherever the Jacobians exist. $B(\cdot)$ is $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere differentiable from Definition 3.1.14 (while $\bar{A}(\cdot)$ — the first $n - \bar{p}$ components of $A(\cdot)$ — is differentiable everywhere). It now remains to show that the final p columns of $\hat{\sigma}(B(\hat{x}))$ lie in the span of the final \bar{p} columns of $J_B(\hat{x})$.

Secondly, for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all \hat{x} , this span is given by the prevailing unspanned state directions $\mathcal{S}^{B(\hat{x})}$. This follows from viewing Equations 3.9 and 3.13: excluding the $\hat{X}_*(\lambda \otimes \mathbb{P})$ -null points where Equation 3.9 does not hold, it follows immediately that the final \bar{p} columns of $J_B(\hat{x})$ are elements of $\mathcal{S}^{B(\hat{x})}$ (given the zero partial derivative in Equation 3.9, these columns conform to Definition 3.1.1 by virtue of Equation 3.13). That these \bar{p} columns' span is equal to $\mathcal{S}^{B(\hat{x})}$ follows from observing that they are linearly independent (from the invertibility of $J_B(\hat{x})$) — we simply need to exclude the points where $\mathcal{S}^{B(\hat{x})}$ has a greater dimension than \bar{p} to ensure that it is not larger than the span of these columns (Definition 3.1.12 ensures that these states are $\hat{X}_*(\lambda \otimes \mathbb{P})$ -null).

Thirdly, the final p columns of $\hat{\sigma}(B(\hat{x}))$ lie in this span. From the proof of Lemma 3.1.4, we have that $\hat{\sigma}(B(\hat{x})) = \sigma(B(\hat{x}))\epsilon(B(\hat{x}))$; without partially spanned factors, the final p columns of $\epsilon(B(\hat{x}))$ are set to lie in $\mathcal{U}^{B(\hat{x})}$ $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere. This causes the final p columns of $\hat{\sigma}(B(\hat{x}))$ to lie in $\mathcal{S}^{B(\hat{x})}$.

Concerning the size and rank of $\sigma^Y(\hat{x})$, first note that, for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all \hat{x} , we have $p = \dim(\mathcal{U}^{B(\hat{x})})$. The discussion below Definition 3.1.2 notes that the unspanned factor directions $\mathcal{U}^{B(\hat{x})}$ either belong to the kernel of the diffusion matrix (which is necessarily a subspace of $\mathcal{U}^{B(\hat{x})}$) or correspond to unspanned state directions (i.e., elements of $\mathcal{S}^{B(\hat{x})}$) that lie in the image of $\sigma(B(\hat{x}))$. To formalise this, define

$$p^{\text{SV}\sigma}(\hat{x}) = \dim\left(\text{Im}(\sigma(B(\hat{x}))) \cap \mathcal{S}^{B(\hat{x})}\right)$$

for all $\hat{x} \in \hat{\mathcal{X}}$, which gives the dimension of the subspace of $\mathcal{U}^{B(\hat{x})}$ that corresponds to non-trivial unspanned state directions. It then follows that, for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all \hat{x} ,

$$p = \text{nul}(\sigma(B(\hat{x}))) + p^{\text{SV}\sigma}(\hat{x}),$$

where $\text{nul}(\cdot)$ gives the dimension of a matrix's kernel. Note that the number of unspanned factor is $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere constant (by the assumption of no partially spanned factors), whereas its subspaces can, in principle, vary across the state space (provided they $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere sum to p). We also have that $\bar{p} = \dim(\mathcal{S}^{B(\hat{x})})$ for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all \hat{x} , which we can similarly decompose:

$$\bar{p} = p^{\text{SV}\sigma}(\hat{x}) + p^{\text{SV}\neg\sigma}(\hat{x}), \tag{3.23}$$

where $p^{\text{SV}\neg\sigma}(\cdot)$ — defined by the above equation — gives the number of dimensions of unspanned integral, up to almost-sure equality (under \mathbb{P} or any equivalent measure). In other words, the process that satisfies the dynamics in the lemma is almost-surely equal to the one that satisfies the dynamics where the final p columns of $\{\hat{\sigma}(\hat{X}_t)\}$ are $\lambda \otimes \mathbb{P}$ -almost everywhere zero.

state directions that do not lie in the image of $\sigma(B(\hat{x}))$; that is,

$$p^{\text{SV}\neg\sigma}(\hat{x}) = \dim \left(\text{Im}(\sigma(B(\hat{x})))^\perp \cap \mathcal{S}^{B(\hat{x})} \right).$$

With regard to the size of the matrix $\sigma^Y(\hat{x})$, consider the difference between its number of rows and number of columns:

$$\begin{aligned} n - \bar{p} - (d - p) &= n - p^{\text{SV}\sigma}(\hat{x}) - p^{\text{SV}\neg\sigma}(\hat{x}) - d + \text{nul}(\sigma(B(\hat{x}))) + p^{\text{SV}\sigma}(\hat{x}) \\ &= n - p^{\text{SV}\neg\sigma}(\hat{x}) - d + \text{nul}(\sigma(B(\hat{x}))) \\ &= n - p^{\text{SV}\neg\sigma}(\hat{x}) - \text{rank}(\sigma(B(\hat{x}))), \end{aligned} \quad (3.24)$$

where the final step uses the rank-nullity theorem. The above must be non-negative (showing there are at least as many rows as columns) because $n - \text{rank}(\sigma(B(\hat{x})))$ is the maximum number of dimensions available for non-trivial vectors in \mathbb{R}^n to lie outside the image of $\sigma(B(\hat{x}))$ (i.e., $n - \text{rank}(\sigma(B(\hat{x})))$ is maximum possible value for $p^{\text{SV}\neg\sigma}(\hat{x})$).

To show that $\sigma^Y(\hat{x})$ has a full rank of $d - p$, first note that

$$\text{rank} \left(J_A(B(\hat{x})) \hat{\sigma}(B(\hat{x})) \right) = \text{rank}(\sigma(B(\hat{x})))$$

for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all \hat{x} (the points where $A(\cdot)$ is not differentiable are excluded), because multiplication by an invertible matrix such as $J_A(B(\hat{x}))$ preserves rank. Using Equation 3.24 and then Equation 3.23, we get

$$\text{rank} \left(J_A(B(\hat{x})) \hat{\sigma}(B(\hat{x})) \right) = d - p + p^{\text{SV}\sigma}(\hat{x}). \quad (3.25)$$

We then claim that, for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all \hat{x} ,

$$\text{rank} \left(J_{\bar{A}}(B(\hat{x})) \hat{\sigma}(B(\hat{x})) \right) = d - p. \quad (3.26)$$

In other words, we claim that the inclusion of the final \bar{p} rows of $J_A(B(\hat{x})) \hat{\sigma}(B(\hat{x}))$ adds only $p^{\text{SV}\sigma}(\hat{x})$ to the rank. To see this, consider expressing the rows of $J_A(B(\hat{x}))$ in terms of a certain basis. These rows themselves form a basis of \mathbb{R}^n , but consider replacing the final \bar{p} rows with a basis for $\mathcal{S}^{B(\hat{x})}$.¹¹ Let this basis for $\mathcal{S}^{B(\hat{x})}$ be given by $p^{\text{SV}\sigma}(\hat{x})$ vectors from

$$\text{Im}(\sigma(B(\hat{x}))) \cap \mathcal{S}^{B(\hat{x})},$$

which is indeed $p^{\text{SV}\sigma}(\hat{x})$ -dimensional (by the definition of $p^{\text{SV}\sigma}(\cdot)$ above), and $\bar{p} - p^{\text{SV}\sigma}(\hat{x}) = p^{\text{SV}\neg\sigma}(\hat{x})$ vectors from

$$\text{Im}(\sigma(B(\hat{x})))^\perp \cap \mathcal{S}^{B(\hat{x})}.$$

Critically, these final $p^{\text{SV}\neg\sigma}(\hat{x})$ vectors lie in the cokernel of $\hat{\sigma}(B(\hat{x}))$ (i.e., the kernel of $\hat{\sigma}(B(\hat{x}))^\top$).¹²

Consider the additional rank that is gained by adding the final \bar{p} rows of $J_A(B(\hat{x})) \hat{\sigma}(B(\hat{x}))$

¹¹This would indeed result in a basis for \mathbb{R}^n , because it was established earlier in the proof that the first $n - \bar{p}$ rows of $J_A(B(\hat{x}))$ — i.e., the rows of $J_{\bar{A}}(B(\hat{x}))$ — are orthogonal to all vectors in $\mathcal{S}^{B(\hat{x})}$.

¹²That the orthogonal complement of a matrix's image is equal to its cokernel is part of the so-called *fundamental theorem of linear algebra* (which also incorporates the rank-nullity theorem) — see Meyer [2000, Ch.4.4] or Strang [1993].

to $J_{\bar{A}}(B(\hat{x}))\hat{\sigma}(B(\hat{x}))$. Using this particular basis, these final rows can be expressed as linear combinations of the first $n - \bar{p}$ rows and the basis of $\mathcal{S}^{B(\hat{x})}$ described above. This reveals that the additional rank cannot exceed $p^{\text{SV}\sigma}(\hat{x})$ — aside from the $p^{\text{SV}\sigma}(\hat{x})$ basis vectors from $\text{Im}(\sigma(B(\hat{x}))) \cap \mathcal{S}^{B(\hat{x})}$, the vectors that feature in any linear combination cannot cause any additional rank (the ones from the first $n - \bar{p}$ rows of $J_{\bar{A}}(B(\hat{x}))\hat{\sigma}(B(\hat{x}))$ can only add dimension that is already accounted for; the ones from the cokernel have no effect after they are multiplied by $\hat{\sigma}(B(\hat{x}))$). This establishes Equation 3.26.

Because we have shown that the final p of the d columns of $J_{\bar{A}}(B(\hat{x}))\hat{\sigma}(B(\hat{x})) = \hat{\hat{\sigma}}(\hat{x})$ are equal to zero ($\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere), the first $d - p$ columns must be linearly independent in order to maintain Equation 3.26. Therefore $\sigma^Y(\hat{x})$ has a full rank of $d - p$, completing the proof. \square

To conclude the section, we illustrate Lemma 3.1.18 by revisiting Examples 2 and 3.

Example 2 (Continued). The \mathcal{S} -transformation suggested above (see Equation 3.11) is \mathcal{C}^2 , so Itô's lemma can be applied to the whole transformed state process:

$$\begin{aligned} d\hat{X}_t &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} dX_t = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -\kappa X_t^{(1)} \\ -\kappa X_t^{(2)} \end{bmatrix} dt + \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} dW_t \\ &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -\kappa X_t^{(1)} \\ -\kappa X_t^{(2)} \end{bmatrix} dt + \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} \bar{\sigma}_1 & \bar{\sigma}_2 \\ \bar{\sigma}_2 & -\bar{\sigma}_1 \end{bmatrix} d\hat{W}_t, \end{aligned} \quad (3.27)$$

which, after simplification and using the inverse $B(\cdot)$ to write the dynamics in term of \hat{X}_t , yields

$$d\hat{X}_t = \begin{bmatrix} -\kappa \hat{X}_t^{(1)} \\ -\kappa \hat{X}_t^{(2)} \end{bmatrix} dt + \begin{bmatrix} \sigma_1 \bar{\sigma}_1 + \sigma_2 \bar{\sigma}_2 & 0 \\ \sigma_1 \bar{\sigma}_1 & \sigma_1 \bar{\sigma}_2 \end{bmatrix} d\hat{W}_t.$$

The spanned component is given by the first state variable ($Y_t = \hat{X}_t^{(1)}$), so that

$$\begin{aligned} dY_t &= -\kappa Y_t dt + \begin{bmatrix} \sigma_1 \bar{\sigma}_1 + \sigma_2 \bar{\sigma}_2 & 0 \end{bmatrix} d\hat{W}_t \\ &= -\kappa Y_t dt + (\sigma_1 \bar{\sigma}_1 + \sigma_2 \bar{\sigma}_2) d\hat{W}_t^{(1)}. \end{aligned} \quad (3.28)$$

The integrand corresponding to the second Brownian motion is equal to zero, as required by Lemma 3.1.18 (recall that there are $p = 1$ unspanned factors in this model).

From Theorem 3.1.9, this TSM is mathematically incomplete — intuitively, the stochastic shocks of the second transformed factor $\{\hat{W}_t^{(2)}\}$ do not enter ZCB price processes, and so any claim involving these shocks will be not be attainable. However, Theorem 3.2.10 below will show the model to be economically complete, and Equation 3.28 gives the primary intuition of this result: $\{Y_t\}$ is adapted to the filtration generated by $\{\hat{W}_t^{(1)}\}$. As bond prices can be given by the spanned component $\{Y_t\}$, the zero-coupon bond filtration $\{\mathcal{F}_t^P\}$ can be bounded in this way. In addition to Lemma 3.1.4, this line of reasoning requires that the spanned component's drift and diffusion functions depend only on the spanned component itself (in other words, that $U_t = \hat{X}_t^{(2)}$ does not appear in the spanned component's dynamics). The proof of Theorem 3.2.10 shows this to be true for the drift function, and, in certain circumstances described in the next section, for the diffusion function.

Example 3 (Continued). Using the \mathcal{S} -transformation suggested above (see Equations 3.15 and 3.16), the dynamics of the spanned component $\{Y_t\}$, which is the first element of the transformed state process $\{\hat{X}_t\}$, can be inferred. Applying Itô's lemma (based on the original dynamics of the

model in Equation 3.7) and simplifying yields

$$dY_t = dA^{(1)}(X_t) = -2\left(\kappa(X_t^{(1)})^2 + \kappa(X_t^{(2)})^2 - \sigma^2\right)dt + 2\sigma \begin{bmatrix} X_t^{(1)} & X_t^{(2)} \end{bmatrix} dW_t.$$

The orthonormal transformation from Lemma 3.1.4 has not yet been applied, so we do not yet have the volatility form of Lemma 3.1.18. A suitable orthonormal transformation $\{\epsilon(X_t)\}$ for this model was described in Equation 3.8 above — applying this, but suppressing some of the calculations, yields

$$\begin{aligned} dY_t &= -2\left(\kappa(X_t^{(1)})^2 + \kappa(X_t^{(2)})^2 - \sigma^2\right)dt + 2\sigma \begin{bmatrix} X_t^{(1)} & X_t^{(2)} \end{bmatrix} \epsilon(X_t) d\hat{W}_t \\ &= -2\left(\kappa(X_t^{(1)})^2 + \kappa(X_t^{(2)})^2 - \sigma^2\right)dt + 2\sigma \begin{bmatrix} \sqrt{(X_t^{(1)})^2 + (X_t^{(2)})^2} & 0 \end{bmatrix} d\hat{W}_t \\ &= 2\kappa\left(\frac{\sigma^2}{\kappa} - Y_t\right)dt + 2\sigma\sqrt{Y_t} d\hat{W}_t^{(1)}. \end{aligned}$$

Like in Example 2, the second transformed factor is not needed for the dynamics of the spanned component.

3.2 Unspanned Factor Roles

An unspanned factor can play a variety of roles; a fully unspanned factor, for instance, could be completely extraneous to a TSM. The goal of this subsection is to examine these various roles and, in particular, to examine the cases where an unspanned factor is responsible for stochastic volatility. We first require the following notion of time- t instantaneous ZCB-covariance.

Definition 3.2.1. For $\tau_1, \tau_2 > 0$, the *bond covariance function* is given, for all $x \in \mathcal{X}$, by

$$v(\tau_1, \tau_2, x) = \nabla P(\tau_1, x)^\top \sigma(x) \sigma(x)^\top \nabla P(\tau_2, x). \quad (3.29)$$

This covariance measure is *instantaneous* in nature because it captures the covariance of ZCB price changes (for ZCBs with times to maturity of τ_1 and τ_2) over the next instant; formally, we have¹³

$$v(\tau_1, \tau_2, X_t) = \lim_{\delta \downarrow 0} \frac{1}{\delta} \mathbb{C}\mathbb{P}^\mathbb{P}[P(\tau_1 - \delta, X_{t+\delta}), P(\tau_2 - \delta, X_{t+\delta}) | \mathcal{F}_t]. \quad (3.30)$$

Given the Brownian framework, this coincides with the time- t rate at which the two ZCB price processes (with maturities $t + \tau_1$ and $t + \tau_2$) are accumulating quadratic covariation. If evaluated at a single time-to-maturity, the bond covariance function reduces to a bond variance (or a bond squared-volatility) function:

$$v(\tau, \tau, X_t) = \nabla P(\tau, X_t)^\top \sigma(X_t) \sigma(X_t)^\top \nabla P(\tau, X_t),$$

which is simply the squared Euclidean norm of the volatility process in the relevant ZCB price process dynamics (as well as the rate at which the ZCB price process is accumulating quadratic variation).

The bond covariance function is a differentiable function of the state process (this follows from Assumptions 2.1.2 and 2.1.5). We have seen, in Section 3.1, that certain movements in the factor

¹³See Crépey [2013, Ch.3.2].

process do not affect the term structure — for any state $x \in \mathcal{X}$, Definition 3.1.2 gives \mathcal{U}^x as the set of these direction vectors. Some of these directions do not affect the bond covariances either, while some of them do; the following definition characterises the latter subset.

Definition 3.2.2. For any $x \in \mathcal{X}$, the *USV factor directions* are given by

$$\mathcal{V}_{\mathcal{U}}^x = \{\delta \in \mathcal{U}^x \mid \nabla v(\tau_1, \tau_2, x)^\top \sigma(x) \delta = 0 \quad \forall \tau_1, \tau_2 > 0\}^\perp,$$

where $\{\cdot\}^\perp$ denotes the orthogonal complement with respect to \mathcal{U}^x .

Example 1 (Continued). The Fong & Vasiček [1991] model contains no non-trivial unspanned factor directions; i.e., $\mathcal{U}^x = \{\mathbf{0}_2\}$ for all $x \in \mathcal{X}$. Therefore, despite the fact that the bond covariance function is sensitive to the state process, no non-trivial vectors are allowed to enter $\mathcal{V}_{\mathcal{U}}^x$ (in other words, despite the stochastic volatility nature of the model, no *unspanned* stochastic volatility directions are exhibited). To illustrate this explicitly, first note that

$$\{\delta \in \mathbb{R}^2 \mid \nabla v(\tau_1, \tau_2, x)^\top \sigma(x) \delta = 0 \quad \forall \tau_1, \tau_2 > 0\} = \{\mathbf{0}_2\}, \quad (3.31)$$

because the bond covariance function is sensitive to both state variables (and the diffusion function straightforwardly links state process and factor movements), so that there are no direction vectors resulting in a directional derivative of zero.¹⁴ Taking the orthogonal complement with respect to \mathbb{R}^2 , we get the factor directions that do affect the TSM's covariance structure:

$$\{\delta \in \mathbb{R}^2 \mid \nabla v(\tau_1, \tau_2, x)^\top \sigma(x) \delta = 0 \quad \forall \tau_1, \tau_2 > 0\}^\perp = \mathbb{R}^2.$$

However, Definition 3.2.2 requires USV factor directions to be unspanned factor directions in the first place — because of the lack of unspanned directions in this example, we have

$$\{\delta \in \mathcal{U}^x \mid \nabla v(\tau_1, \tau_2, x)^\top \sigma(x) \delta = 0 \quad \forall \tau_1, \tau_2 > 0\} = \{\mathbf{0}_2\},$$

so that when the orthogonal complement *with respect to* \mathcal{U}^x is taken, we have

$$\mathcal{V}_{\mathcal{U}}^x = \{\delta \in \mathcal{U}^x \mid \nabla v(\tau_1, \tau_2, x)^\top \sigma(x) \delta = 0 \quad \forall \tau_1, \tau_2 > 0\}^\perp = \{\mathbf{0}_2\}.$$

In the general case, if $\delta \in \mathcal{V}_{\mathcal{U}}^x$, δ exhibits two properties: it is *unspanned* ($\delta \in \mathcal{U}^x$), and δ -directional movements of the factor process affect state- x bond covariances (in order to belong to the complement, there must exist some $\tau_1^*, \tau_2^* > 0$ such that $\nabla v(\tau_1^*, \tau_2^*, x)^\top \sigma(x) \delta \neq 0$). We term such directions *unspanned stochastic volatility (USV) factor directions*.

More generally, note why the bond covariance function is a suitable notion of covariance to use in the definition of USV factor directions: if an unspanned movement in the factor process is to affect the covariance of any aspects of the future term structure, *it must do so by affecting the covariance experienced in the ensuing instant* (which, from Equation 3.30, is captured by the bond covariance function).

¹⁴The Fong & Vasiček [1991] model is introduced by Equations 1.6 and 1.7, after which it is mentioned that the Brownian motions can be correlated. In the introduction of Example 1, we take $\{W_t^{(1)}\}$ and $\{W_t^{(2)}\}$ to constitute a standard, uncorrelated \mathbb{Q} -Brownian motion; a correlation can nevertheless be reflected by setting

$$\sigma(X_t) = \sigma(r_t, v_t) = \begin{bmatrix} \sqrt{v_t} & 0 \\ \rho \sigma_v \sqrt{v_t} & \sqrt{1 - \rho^2} \sigma_v \sqrt{v_t} \end{bmatrix}.$$

Provided that $-1 < \rho < 1$, the statement that the two dimensions of factor movements are linked to the two dimensions of state process movements is accurate. If $\rho \in \{1, -1\}$, then the second factor $\{W_t^{(2)}\}$ would not affect bond prices or covariances, and the set given in Equation 3.31 would be one-dimensional.

Also note that Definition 3.2.2 can be said to capture volatility-affecting directions, in a broad sense of volatility: only some aspect of the whole ZCB instantaneous *covariance* structure need be affected for a direction to enter $\mathcal{V}_{\mathcal{U}}^x$. One could distinguish between directions affecting (squared) volatility (where the τ_1^* and τ_2^* discussed above coincide) and directions affecting bond covariances for distinct maturities. In other words, one could think of the USV factor directions $\mathcal{V}_{\mathcal{U}}^x$ as including both USV directions in the strict sense of volatility, as well as *unspanned stochastic correlation* directions. Beginning with the following definition, we characterise unspanned stochastic volatility without this distinction. In Section 3.4.4, we discuss strict USV versus unspanned stochastic correlation.

As in Section 3.1, it is the dimension of $\mathcal{V}_{\mathcal{U}}^x$ over the state space that is of interest.

Definition 3.2.3. The $\mathcal{V}_{\mathcal{U}}$ -partition of the state space is given by

$$\mathcal{X}_{\mathcal{V},\mathcal{U}}^{(i)} = \{x \in \mathcal{X} \mid \dim(\mathcal{V}_{\mathcal{U}}^x) = i\}$$

for $i = 0, 1, \dots, d$.

Lemma 3.1.4 shows that after an orthonormal factor transformation, ZCBs are not sensitive to the stochastic shocks of the final i transformed factors when $X_t \in \mathcal{X}_{\mathcal{U}}^{(i)}$ (that is, when \mathcal{U}^x is i -dimensional). The transformation is based on some orthonormal basis of \mathbb{R}^d for all $x \in \mathcal{X}$, where the final i basis vectors form a basis for \mathcal{U}^x . One could specifically choose this basis such that its final few vectors form a basis for $\mathcal{V}_{\mathcal{U}}^x$ (in other words, when choosing basis vectors for \mathcal{U}^x , one can start by choosing a basis for the linear subspace $\mathcal{V}_{\mathcal{U}}^x$). Then, when $X_t \in \mathcal{X}_{\mathcal{V},\mathcal{U}}^{(j)}$, the stochastic shocks of the final j transformed factors *do not feed into ZCB prices, but do feed into ZCB covariances*. Such a construction of the orthonormal basis function $\epsilon(\cdot)$ gives the following corollary to Lemma 3.1.4.

Corollary 3.2.4. There exists an orthonormal factor process transformation such that ZCB dynamics can be given in terms of the transformed factor process $\{\hat{W}_t\}$ as in Lemma 3.1.4, namely

$$dP_{tT} = r_t P_{tT} dt + \nabla P(T-t, X_t)^\top \hat{\sigma}(X_t) d\hat{W}_t,$$

or

$$dP_{tT} = (r_t P_{tT} + \nabla P(T-t, X_t)^\top \sigma(X_t) \zeta_t) dt + \nabla P(T-t, X_t)^\top \hat{\sigma}(X_t) d\hat{W}_t^\mathbb{P},$$

and such that, for $i = 1, 2, \dots, d$,

$$[\nabla P(\tau, x)^\top \hat{\sigma}(x)]_i = 0 \quad \forall \tau > 0, \quad \text{and} \quad [\nabla v(\tau_1^*, \tau_2^*, x)^\top \hat{\sigma}(x)]_i \neq 0 \quad \text{for some } \tau_1^*, \tau_2^* > 0,$$

for all $x \in \bigcup_{j=d-i+1}^d \mathcal{X}_{\mathcal{V},\mathcal{U}}^{(j)}$.

This follows from the above discussion of how the basis vectors are selected, and, given that selection, from observing that if $x \in \mathcal{X}_{\mathcal{V},\mathcal{U}}^{(i)}$, the final i elements of

$$\nabla v(\tau_1, \tau_2, x)^\top \hat{\sigma}(x) = \nabla v(\tau_1, \tau_2, x)^\top \sigma(x) \epsilon(x)$$

are *not* equal to zero for all $\tau_1, \tau_2 > 0$ (this follows from Definition 3.2.2). This implies the existence of the τ_1^* and τ_2^* used in the corollary.

In summary, if $X_t \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{V},\mathcal{U}}^{(j)}$, then the stochastic shocks of at least one (transformed) factor do not enter ZCB price processes (and therefore cannot be hedged by holding ZCBs), but do

enter (at least some) bond covariances. This is the essence of the USV concept. If $\bigcup_{j=1}^d \mathcal{X}_{\mathcal{V}, \mathcal{U}}^{(j)}$ is non-empty and $X_*(\lambda \otimes \mathbb{P})$ -non-null, then there is some probability that this state of affairs will occur and *persist* (for a λ -non-null period of time), which is the motivation behind the following definition.

Definition 3.2.5. Let r equal the largest value of i for which $\mathcal{X}_{\mathcal{V}, \mathcal{U}}^{(i)}$ is $X_*(\lambda \otimes \mathbb{P})$ -non-null. The TSM in question is said to have r *USV factors*.

TSMs exhibiting a USV factor are analogous to the [Heston \[1993\]](#) model, as discussed in Section 1.2: they exhibit stochastic volatility risk (for TSMs with a USV factor, this is the risk inherent in the random changes to the final few elements of $\{\hat{W}_t\}$) that is unobtainable via the fundamental assets (which are the ZCBs in the case of TSMs).

Example 4. In Section 1.2, a simple USV model was briefly introduced. For a full specification of the model, consider the following two-dimensional state process dynamics:

$$dX_t = d \begin{bmatrix} \pi_t \\ v_t \end{bmatrix} = \begin{bmatrix} \kappa_\pi(\theta_\pi - \pi_t) \\ \kappa_v(\theta_v - v_t) \end{bmatrix} dt + \begin{bmatrix} v_t(\pi_t - \underline{\pi}) & 0 \\ 0 & \sigma_v \sqrt{v_t} \end{bmatrix} dW_t^\mathbb{P},$$

where $\kappa_\pi, \kappa_v, \theta_\pi, \theta_v, \underline{\pi}$ and σ_v are positive constants and $\{W_t^\mathbb{P}\}$ is a two-dimensional \mathbb{P} -Brownian motion. Provided that $\pi_0 > 0$ and that $\theta_\pi > \underline{\pi} > 0$, $\{\pi_t\}$ remains strictly positive and is suitable to be taken as a pricing kernel process.¹⁵ Based on the rational term structure model pricing expression in Equation 1.11, the model's bond pricing function is

$$P(\tau, x) = \frac{\theta_\pi + e^{-\kappa_\pi \tau} (x^{(1)} - \theta_\pi)}{x^{(1)}}.$$

Because the second state variable does not affect the term structure (and because of the diagonal diffusion matrix), we have that

$$\mathcal{U}^x = \left\{ \begin{bmatrix} 0 & \alpha \end{bmatrix}^\top \mid \alpha \in \mathbb{R} \right\}.$$

This is everywhere one-dimensional; the TSM thus exhibits one unspanned factor.¹⁶ The bond covariance function is given by

$$v(\tau_1, \tau_2, x) = \frac{\partial P(\tau_1, x)}{\partial x^{(1)}} \frac{\partial P(\tau_2, x)}{\partial x^{(1)}} (x^{(1)} - \underline{\pi})^2 (x^{(2)})^2. \quad (3.32)$$

The fact that the bond covariance depends on the second state variable is the intuitive reason that a USV factor is exhibited by this TSM. To confirm this, note that

$$\nabla v(\tau_1, \tau_2, x)^\top = \begin{bmatrix} 0 & \frac{\partial P(\tau_1, x)}{\partial x^{(1)}} \frac{\partial P(\tau_2, x)}{\partial x^{(1)}} (x^{(1)} - \underline{\pi})^2 2x^{(2)} \end{bmatrix}.$$

If one ignores $X_*(\lambda \otimes \mathbb{P})$ -null states where $x^{(1)} = \underline{\pi}$ or $x^{(2)} = 0$, it follows that $\nabla v(\tau_1, \tau_2, x)^\top \sigma(x) \delta \neq 0$ for all $\delta \in \mathcal{U}^x$.¹⁷ For $X_*(\lambda \otimes \mathbb{P})$ -almost all states x , we thus have that $\mathcal{V}_{\mathcal{U}}^x = \mathcal{U}^x$, so that $\mathcal{X}_{\mathcal{V}, \mathcal{U}}^{(1)}$ covers $X_*(\lambda \otimes \mathbb{P})$ -almost all of the state space. From Definition 3.2.5, the TSM exhibits one USV

¹⁵If $\pi_t = \underline{\pi}$, then $\{\pi_t\}$ has no local volatility but does have a positive mean reversion. Because the sample paths are continuous, the process $\{\pi_t\}$ cannot therefore cross the lower bound of $\underline{\pi}$. A closely related mechanism of maintaining positivity is used in Section 5.2.1.

¹⁶It also exhibits one unspanned state variable (as $\mathcal{S}^x = \mathcal{U}^x$ $X_*(\lambda \otimes \mathbb{P})$ -almost everywhere). It is noteworthy that the TSM is already in an \mathcal{S} -transformed form; Equation 3.9 in Definition 3.1.14 is satisfied by the original, untransformed bond pricing function.

¹⁷This holds for any τ_1 and τ_2 . It would suffice for it hold for some pair τ_1^* and τ_2^* , which is what Corollary 3.2.4 specifies, in effect, for a general TSM with a USV factor.

factor. In this particular model, a factor process transformation is not necessary for Corollary 3.2.4 to apply — we have that

$$\begin{aligned} dP_{tT} &= (r_t P_{tT} + \nabla P(T-t, X_t)^\top \sigma(X_t) \zeta_t) dt + \nabla P(T-t, X_t)^\top \sigma(X_t) dW_t^\mathbb{P} \\ &= (r_t P_{tT} + \nabla P(T-t, X_t)^\top \sigma(X_t) \zeta_t) dt + \left[v_t(\pi_t - \underline{\pi}) \frac{\partial P(T-t, x)}{\partial x^{(1)}} \Big|_{x=X_t} \quad 0 \right] dW_t^\mathbb{P}, \end{aligned}$$

so that the second (untransformed) factor is not an integrator for the ZCB price processes, even though its changes do affect ZCB volatility (the directions in \mathcal{V}_U^x correspond exactly to this second factor). Note that the market-price-of-risk process $\{\zeta_t\}$ is implicit in the specification of the pricing kernel process; this needs to be deduced if the model's risk-neutral measure is to be explicitly identified.

Theorem 3.1.9 above shows that unspanned factors cause mathematical incompleteness, as per Definition 2.2.5. We establish below (in Theorem 3.2.10) that unspanned factors do not necessarily cause economic incompleteness (Definition 2.2.6), but we claim that USV factors do necessarily have this effect. For this result, in addition to Assumption 3.1.6, we require an assumption of some regularity of the function $v(\tau_1^*, \tau_2^*, \cdot)$ (where τ_1^* and τ_2^* are the pair of maturities given by Corollary 3.2.4). Also, recall the exclusion of partially spanned factors explained in Remark 3.1.16.

Theorem 3.2.6. Given Assumptions 3.1.6 and 3.3.4, if a TSM has any USV factors and also has no partially spanned factors, then it is economically incomplete.

The proof is given in Section 3.3.2, as is Assumption 3.3.4. Our (partial) converse to this result will require the notion of *USV state variables*. This notion is developed in the way that the definition of unspanned state variables followed that of unspanned factors in Section 3.1.

Definition 3.2.7. For any $x \in \mathcal{X}$, the *USV state directions* are given by

$$\mathcal{V}_S^x = \{\delta \in \mathcal{S}^x \mid \nabla v(\tau_1, \tau_2, x)^\top \delta = 0 \quad \forall \tau_1, \tau_2 > 0\}^\perp,$$

where $\{\cdot\}^\perp$ denotes the orthogonal complement with respect to \mathcal{S}^x .

Definition 3.2.8. The \mathcal{V}_S -*partition* of the state space is given by

$$\mathcal{X}_{\mathcal{V}, \mathcal{S}}^{(i)} = \{x \in \mathcal{X} \mid \dim(\mathcal{V}_S^x) = i\}$$

for $i = 0, 1, \dots, n$.

Definition 3.2.9. Let \bar{r} equal the largest value of i for which $\mathcal{X}_{\mathcal{V}, \mathcal{S}}^{(i)}$ is $\lambda \otimes \mathbb{P}$ -non-null. The TSM in question is said to have \bar{r} *USV state variables*.

The converse result to Theorem 3.2.6 would be that an economically incomplete TSM must exhibit a USV factor, or, in terms of the contrapositive, that a TSM without USV factors is economically complete. Our partial converse requires some stronger conditions, given in the following.

Theorem 3.2.10. Given Assumptions 3.1.6 and 3.3.8, if a TSM has no partially spanned factors or state variables, has no USV state variables and admits an \mathcal{S} -transformation, then it is economically complete.

The proof is given in Section 3.3.3, as is Assumption 3.3.8, which is an exact analogue of Assumption 3.3.2 (Assumption 3.3.2 was used in Theorem 3.1.9 and is discussed there — see the first subsection of Section 3.3.1, as well as Remark 3.3.9, for further discussion). Note the

stronger state process conditions needed in Theorem 3.2.10 compared to Theorem 3.2.6: an \mathcal{S} -transformation and the absence of partially spanned state variables are required. The former is discussed in Remark 3.1.15 and the latter in Remark 3.1.16 (where it is noted that partially spanned factors and state variables do not appear in the literature, and their discussion is postponed to Section 3.4.1). Finally, note that the absence of USV state variables is a stronger condition than the absence of USV factors; this difference, along with the question of what constitutes a so-called *USV model*, is addressed in Remark 3.2.12 below.

Critically, we see that unspanned factors or state variables, though they cause mathematical incompleteness, are not necessarily an obstacle to economic completeness; this depends on the particular role that they play. If they do not affect the covariance structure of the TSM, then — modulo the requirements listed above, which tend to be met in practice — they do not induce economic completeness.

More elaborate models exhibiting USV factors and state variables appear in the final subsection of Section 4.1.2 (where they are used to simulate data that benchmark the empirical results of Chapter 4). There we give a remark about the modelling of USV, addressing the fact that unspanned factors are a prerequisite to USV factors. The Fong & Vasíček [1991] model in Example 1 shows that, although constructing TSMs with stochastic volatility is relatively easy, ensuring that the relevant state variables are unspanned is difficult. Constructing USV models therefore requires a modelling framework that is somehow conducive to unspanned factors and state variables — this point is discussed in Section 4.1.2, specifically in Remark 4.1.1.

The two following remarks conclude the section.

Remark 3.2.11. This section has described a particular role that unspanned factors or state variables can play — despite having no impact on the term structure, they might impact the volatility of future bond prices and yields. If they do not, they may be truly redundant to a model. In Examples 2 and 3, the unspanned factor and state variable in either model do not give rise to USV factors or state variables — the state processes’ constant diffusion matrices cause \mathcal{V}_U^x and \mathcal{V}_S^x to be trivial sets. Theorem 3.2.10 implies that these models are economically complete. The unspanned state variable in each model appears to play no role beside expanding the set of claims that need to be attainable for mathematical completeness — it appears that each model can be fully described by the spanned state variable, based on a single Brownian motion (in Example 2, for instance, the spanned state variable can be defined by Equation 3.28, and bonds given by the reduced bond pricing function in Equation 3.12 — this gives a state-process-bond-pricing-function pair that was considered the essence of a TSM in Section 2.1).

However, we note that there is one potential role that unspanned state variables and factors can play aside from inducing USV, which is to be involved in the market-price-of-risk process $\{\zeta_t\}$ relating the real-world and risk-neutral Brownian motions (see Equation 2.3). For example, ζ_t could be a function of the unspanned component of the state process U_t , in which case the unspanned state variables have an effect on the distribution of future ZCB prices *under the real-world measure* \mathbb{P} . Equation 3.28 in Example 2 would then *not* fully describe the spanned state variable, in that its real-world dynamics (and therefore the real-world dynamics of ZCBs) depend on the unspanned state variable, and in turn on $\{\hat{W}_t^{(2)}\}$. State variables of this sort are studied by Duffee [2011] and Joslin, Priebsch & Singleton [2014], and their modelling is addressed by Filipović et al. [2017].

Remark 3.2.12. This section has formally defined USV factors and USV state variables, while the literature refers more generally to *models exhibiting USV* or *USV models*. If a model exhibits at least one USV factor (Definition 3.2.5), then it should certainly be said to exhibit USV — as per the discussion below Corollary 3.2.4, a USV factor entails stochastic shocks appearing in ZCB

volatility but not in ZCB prices, which is the essence of USV.

If a TSM exhibits a USV state variable but not a USV factor, the situation is less clear. This means that (over some non-null part of the state space and for some τ_1^* and τ_2^*) $\nabla v(\tau_1^*, \tau_2^*, x)^\top \delta \neq 0$ for some unspanned state direction $\delta \in \mathcal{S}^x$, but also that such directions do not lie in the image of $\sigma(x)$ (if they did, a USV factor would straightforwardly exist). In qualitative terms, there are state process movements that do not affect the term structure, that do affect bond covariance, but *are not directly caused by the stochastic shocks of factor process* (as factor shocks are transmitted to the state process via $\sigma(x)$). To highlight this, we need a slightly more demanding version of the \mathcal{S} -transformation than that given by Definition 3.1.14. Assuming the absence of partially spanned state variables, an \mathcal{S} -transformation requires that the final \bar{p} columns (where \bar{p} is the number of unspanned state variables) of the inverse Jacobian $J_B(\hat{x})$ be elements of $\mathcal{S}^{B(\hat{x})}$ (see the discussion below Equation 3.13), or, equivalently, that the first $n - \bar{p}$ rows of $J_A(x)$ be elements of $(\mathcal{S}^x)^\perp$. Here we also require that as many of the remaining rows of $J_A(x)$ as possible be elements of $\text{Im}(\sigma(x))^\perp$. This is always possible for at least one row; given the assumption of USV state variables without USV factors, it follows that

$$\dim \left(\text{Im}(\sigma(x))^\perp \cap \mathcal{S}^x \right) \geq 1. \quad (3.33)$$

Note that in the proof of Lemma 3.1.18, this dimension was effectively denoted $p^{\text{SV}-\sigma}(A(x))$. These rows of $J_A(x)$ are then part of the cokernel of $\sigma(x)$ (see Footnote 12). Assuming that $A^{(k)}(\cdot)$ (where k corresponds to one of these rows) is \mathcal{C}^2 , we therefore have the following with Itô's lemma:

$$\begin{aligned} d\hat{X}_t^{(k)} &= \left(J_{A^{(k)}}(X_t) \mu(X_t) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 A^{(k)}(x)}{\partial x^{(i)} \partial x^{(j)}} \Big|_{x=X_t} [\sigma(X_t) \sigma(X_t)^\top]_{ij} \right) dt + J_{A^{(k)}}(X_t) \sigma(X_t) dW_t \\ &= \left(J_{A^{(k)}}(X_t) \mu(X_t) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{\partial^2 A^{(k)}(x)}{\partial x^{(i)} \partial x^{(j)}} \Big|_{x=X_t} [\sigma(X_t) \sigma(X_t)^\top]_{ij} \right) dt + \mathbf{0}_d^\top dW_t \\ &= \mu^{Y,(k)}(\hat{X}_t) dt, \end{aligned}$$

where $\mu^{Y,(k)}(\cdot)$ denotes the k th component of $\mu^Y(\cdot)$, which is defined in Equation 3.21. This illustrates the nature of USV state variables that are unaccompanied by USV factors — provided an \mathcal{S} -transformation can make the underlying USV direction dimensions correspond to specific state variables, these state variables are *locally deterministic*, and the lack of stochastic shock transmission is explicit.

Now consider exactly how $\mu^{Y,(k)}(\cdot)$ depends on the transformed state process. If — for all values of k corresponding to these locally deterministic state variables — $\mu^{Y,(k)}(\cdot)$ depends only on the spanned component, or on other locally deterministic state variables, then the model does not exhibit USV in any important sense. Indeed, Theorem 3.2.10 can be extended to capture these cases — confirmation given in Remark 3.3.14 in Section 3.3.3. If, however, $\mu^{Y,(k)}(\cdot)$ depends on other parts of the unspanned component (not only itself $X_t^{(k)}$, or the other locally deterministic $X_t^{(j)}$), then the TSM exhibits USV in a weak sense: ZCB volatility can change in a way that is unrelated to the term structure, but this change is *smooth*, in that it does not involve any quadratic variation from the underlying factors.

3.3 Proofs of Completeness Results

Note that the lemmas in this section do not explicitly state the assumed conditions of the theorem in question.

3.3.1 Proof of Theorem 3.1.9

The proof is organised into three subsections, the first two of which prove that lack of unspanned factors implies mathematical completeness, and the last of which proves the converse.

Mathematical Completeness I: Invertible Matrix Process

The first aspect of the proof assumes the absence of unspanned factors. From Definition 3.1.7, this implies that $\bigcup_{j=1}^d \mathcal{X}_U^{(j)}$ is $X_*(\lambda \otimes \mathbb{P})$ -null.

In Section 2.2, a $d \times n$ bond-sensitivity matrix process $\{\nabla P_t^\mathcal{T}\}$ is defined from a (piecewise-constant) maturity process $\{\mathcal{T}_t\}$. In the time-homogeneous context of Chapter 3 (see Assumption 3.0.1), this definition amounts to

$$[\nabla P_t^\mathcal{T}]_{ij} = \frac{\partial P([\mathcal{T}_t]_i - t, x)}{\partial x^{(j)}} \Big|_{x=X_t}.$$

The product of a bond-sensitivity matrix process and the diffusion process of the TSM gives a $d \times d$ matrix process $\{\nabla P_t^\mathcal{T} \sigma(X_t)\}$. Completeness of the TSM requires that this matrix process, for some underlying maturity process $\{\mathcal{T}_t\}$, is $\lambda \otimes \mathbb{P}$ -almost surely invertible — this matrix process constitutes part of the diffusion process of the corresponding ZCB portfolio (given by Definition 2.2.1), and invertibility (up to $\lambda \otimes \mathbb{P}$ -nullness) is needed if the diffusion process of a general claim is to be matched (these details appear in the second subsection of the proof; this subsection addresses only the invertibility of $\{\nabla P_t^\mathcal{T} \sigma(X_t)\}$).

In principle, the ability to find suitable maturities follows straightforwardly from the assumed lack of unspanned factors: at any time t and for any state X_t , we claim that d maturity times $\mathcal{T}^* = \{T_i\}$ can be selected such that associated matrix $\nabla P_t^{\mathcal{T}^*} \sigma(X_t)$ is invertible (at t), provided that X_t is not an element of the higher dimensional blocks of the \mathcal{U} -partition $\bigcup_{j=1}^d \mathcal{X}_U^{(j)}$. This claim is given in Lemma 3.3.1 below. Because the excluded subset $\bigcup_{j=1}^d \mathcal{X}_U^{(j)}$ is by assumption $X_*(\lambda \otimes \mathbb{P})$ -null, there is no probability that the state process occupies it for long enough to prevent $\lambda \otimes \mathbb{P}$ -almost surely invertibility.

Note, however, that this involves *selecting maturity times separately for each time point t* , whereas the maturity process of a ZCB portfolio needs to be piecewise constant and thus *selected in advance for an interval of time points*. This introduces two potential problems. Firstly, although the subset $\bigcup_{j=1}^d \mathcal{X}_U^{(j)}$ is itself $X_*(\lambda \otimes \mathbb{P})$ -null, and so cannot be occupied in a material way, one can face problems *around* this subset. For instance, if $X_0 \in \bigcup_{j=1}^d \mathcal{X}_U^{(j)}$, there is no probability of the state remaining inside $\bigcup_{j=1}^d \mathcal{X}_U^{(j)}$ after $t = 0$, but one may not be able to select maturities at $t = 0$ such that $\{\nabla P_t^\mathcal{T} \sigma(X_t)\}$ becomes invertible *immediately after* $t = 0$ with probability one. Secondly, while one can always choose maturities at a particular time point t (assuming $X_t \in \mathcal{X}_U^{(0)}$) that give invertibility, one cannot be certain how long this invertibility will last.

We now prove the above-mentioned lemma, and then explain why these two problems have little practical relevance.

Lemma 3.3.1. Assuming $X_t \in \mathcal{X}_U^{(0)}$, there exists a d -dimensional maturity process $\{\mathcal{T}_t\}$ such that $\nabla P_t^\mathcal{T} \sigma(X_t)$ is invertible.

Proof. Suppose that suitable maturities cannot be identified; suppose that only $\bar{d} < d$ maturities $\tilde{\mathcal{T}} = \{T_i\}_{i=1}^{\bar{d}}$ can be found that result in the corresponding $\bar{d} \times d$ matrix $\nabla P_t^{\tilde{\mathcal{T}}} \sigma(X_t)$ having a full rank of \bar{d} , but that further maturities that add rows to *and* increase the rank of $\nabla P_t^{\tilde{\mathcal{T}}} \sigma(X_t)$ cannot be found. As a consequence of the rank-nullity theorem, $\nabla P_t^{\tilde{\mathcal{T}}} \sigma(X_t)$ has a non-trivial kernel (because it has a rank lower than number of columns); that is, there exists non-zero $\delta \in \mathbb{R}^d$ such that $\nabla P_t^{\tilde{\mathcal{T}}} \sigma(X_t) \delta = 0$. This shows that $\nabla P(\tau, X_t)^\top \sigma(X_t) \delta = 0$ for all $\tau \in \tilde{\mathcal{T}}$. Because no other maturities choices were able to increase the rank, it also follows that $\nabla P(\tau, X_t)^\top \sigma(X_t) \delta = 0$ for all $\tau > 0$, and therefore that $\delta \in \mathcal{U}^{X_t}$.¹⁸ With δ being non-zero, this shows that \mathcal{U}^{X_t} is at least one-dimensional, contradicting the assumption that $X_t \in \mathcal{X}_\mathcal{U}^{(0)}$. One therefore can find d maturities $\mathcal{T} = \{T_i\}$ such that $\nabla P_t^{\mathcal{T}} \sigma(X_t)$ is invertible. \square

Lemma 3.3.1 guarantees the existence of suitable maturities at any particular time point. Importantly though, for TSMs *in practice*, one finds that *any set of d distinct maturities are suitable* in this sense of leading to an invertible $\nabla P_t^{\mathcal{T}} \sigma(X_t)$. We formalise this in an assumption of *general maturity availability*.

Assumption 3.3.2. If $X_t \in \mathcal{X}_\mathcal{U}^{(0)}$, any d -dimensional maturity process $\{\mathcal{T}_t\}$, provided that the d maturities \mathcal{T}_t are distinct, gives rise to an invertible matrix $\nabla P_t^{\mathcal{T}} \sigma(X_t)$.

This assumption is discussed further in Remark 3.3.9 in Section 3.3.3. The author is not aware of a TSM that violates it.

Note that Assumption 3.3.2 is sufficient to address the two problems mentioned above. There is no problem around the higher-dimensional blocks $\bigcup_{j=1}^d \mathcal{X}_\mathcal{U}^{(j)}$, because *any* ZCB selection gives invertibility outside of the $X_*(\lambda \otimes \mathbb{P})$ -null $\bigcup_{j=1}^d \mathcal{X}_\mathcal{U}^{(j)}$, and so $\lambda \otimes \mathbb{P}$ -almost sure invertibility follows. The problem of a certain set of maturities not sustaining invertibility is also precluded — once initial maturities are selected, one can ignore the $\lambda \otimes \mathbb{P}$ -null visits to $\bigcup_{j=1}^d \mathcal{X}_\mathcal{U}^{(j)}$ and simply wait for the maturities to be reached (whereas, in the general case, the matrix may become singular before this), at which point one selects new maturities that become the second part of the piecewise-constant maturity process. Assuming one does not choose maturities that are increasingly close to the calendar time, maturities can be reselected in this way a finite number of times until the horizon S is reached.

Assumption 3.3.2 is thus sufficient for the existence of an $\lambda \otimes \mathbb{P}$ -almost surely invertible process $\{\nabla P_t^{\mathcal{T}^*} \sigma(X_t)\}$, which is the goal of this subsection. While Theorem 3.1.9 depends on Assumption 3.3.2, this could be relaxed if one included assumptions that address the two above-mentioned problems. This is explored in Appendix B.

Remark 3.3.3. In Remark 3.1.10, our accounting for *state-dependent* unspanned directions is noted. The need for this can be seen in the requirement of an invertible matrix process. The proof of Lemma 3.3.1 shows that an inability to find suitable maturities is equivalent to the existence of a non-trivial unspanned factor direction δ .¹⁹ Whether these directions change over the life of the TSM does not affect this equivalence — non-trivial unspanned factor directions preclude an invertible $\nabla P_t^{\mathcal{T}^*} \sigma(X_t)$, whether or not these directions move with the state process.

¹⁸This is because if any τ^* existed such that $\nabla P(\tau^*, X_t)^\top \sigma(X_t) \delta \neq 0$, then τ^* could have been added to $\tilde{\mathcal{T}}$ and would prevent δ from remaining in the new $(\bar{d} + 1) \times d$ matrix's kernel. This reduction in the nullity implies an increase in the rank, which we supposed to be impossible.

¹⁹The proof shows that an inability to find suitable maturities implies the existence of a non-trivial unspanned factor direction. The converse is straightforwardly true — non-trivial unspanned factor directions imply that any candidate matrix $\nabla P_t^{\mathcal{T}} \sigma(X_t)$ has a non-trivial kernel and is therefore singular.

Mathematical Completeness II: Attainability of a General Claim

Consider now a general T^* -claim Θ that satisfies the boundedness condition in Equation 2.11, ensuring that

$$M_t = \mathbb{E}^{\mathbb{Q}}[D_{T^*}\Theta \mid \mathcal{F}_t]$$

is well-defined for all $t \in [0, T^*]$. The process $\{M_t\}$ is clearly a $(\mathbb{Q}, \{\mathcal{F}_t\})$ -martingale. The martingale representation theorem asserts that there exists an adapted, d -dimensional process $\{\Gamma_t\}$ such that²⁰

$$M_t = M_0 + \int_0^t \Gamma_u^\top dW_u$$

for all $t \in [0, T^*]$. Consider a maturity process $\{\mathcal{T}_t\}$ that results in an $\lambda \otimes \mathbb{P}$ -almost surely invertible $\{\nabla P_t^\top \sigma(X_t)\}$, as per the previous subsection. Suppose we construct a portfolio based on $\{\mathcal{T}_t\}$ and on a holdings process $\{\phi_t\}$ defined with

$$\phi_t = (\nabla P_t^\top \sigma(X_t))^{-\top} \frac{\Gamma_t}{D_t}.$$

Defining the resultant portfolio's value process $\{V_t^{\mathcal{T}, \phi}\}$ as per Definition 2.2.1, its discounted dynamics are given by Equation 2.8 and can be simplified thus:²¹

$$\begin{aligned} d(D_t V_t^{\mathcal{T}, \phi}) &= D_t \phi_t^\top \nabla P_t^\top \sigma(X_t) dW_t, \\ &= \frac{D_t \Gamma_t^\top}{D_t} (\nabla P_t^\top \sigma(X_t))^{-1} \nabla P_t^\top \sigma(X_t) dW_t, \\ &= \Gamma_t^\top dW_t. \end{aligned}$$

If the initial capital of the portfolio $V_0^{\mathcal{T}, \phi}$ is set to M_0 , it follows that $D_{T^*} V_{T^*}^{\mathcal{T}, \phi} = M_{T^*}$. Noting that $M_{T^*} = D_{T^*} \Theta$ (because both Θ and D_{T^*} are \mathcal{F}_{T^*} -measurable), we have that $V_{T^*}^{\mathcal{T}, \phi} = \Theta$. This shows that our definition of mathematical completeness (Definition 2.2.5) is satisfied, completing the first half of the proof.

Lack of Unspanned Factors from Mathematical Completeness

To prove the converse — i.e., that mathematical completeness of a TSM implies that it lacks unspanned factors — consider the following S -claim:

$$\Theta^* = \frac{1}{D_S} \left(\int_0^S \left[(\mathbf{0}_{d-1})^\top \quad \mathbb{I}_{X_u \in \bigcup_{j=1}^d \mathcal{X}_u^{(j)}} \right] \epsilon(X_u)^{-1} dW_u \right),$$

where $\epsilon(\cdot)$ is the orthonormal basis described in Lemma 3.1.4, so that

$$\Theta^* = \frac{1}{D_S} \left(\int_0^S \left[(\mathbf{0}_{d-1})^\top \quad \mathbb{I}_{X_u \in \bigcup_{j=1}^d \mathcal{X}_u^{(j)}} \right] d\hat{W}_u \right). \quad (3.34)$$

²⁰See Filipović [2009, Ch.4.4] for a statement of the martingale representation theorem, as well as a derivation of the extension we have used here: although the filtration $\{\mathcal{F}_t\}$ is generated by $\{W_t^\mathbb{P}\}$, a representation integrand exists for a martingale under any equivalent probability measure. Shreve [2004, Ch.5.3] also gives this result, as a corollary to the martingale representation theorem.

²¹One can safely ignore the paths and time points for which $\{\nabla P_t^\top \sigma(X_t)\}$ is non-invertible — see Footnote 10.

Note that Θ^* is \mathcal{F}_S -measurable and also satisfies Equation 2.11, making its attainability necessary for mathematical completeness.²² We assume mathematical completeness, and therefore the attainability of Θ^* , and show that this implies the absence of unspanned factors. Recalling that Lemma 3.1.4 makes unspanned factors correspond to the final elements of the transformed factor process, the construction of Θ^* ensures that it involves an unspanned factor, if any are in fact present, in an explicit way. More generally, any claim involving unspanned factors will be problematic to hedge.

Let $\{\mathcal{T}_t\}$ and $\{\phi_t\}$ denote the maturity and holdings processes of the (admissible) portfolio value process $\{V_t^{\mathcal{T},\phi}\}$ that attains Θ^* , implying that

$$D_S V_S^{\mathcal{T},\phi} = D_S \Theta^*.$$

Taking an \mathcal{F}_t -conditional \mathbb{Q} -expectation on either side, and noting that the integrand in Equation 3.34 is bounded between zero and one, we get

$$D_t V_t^{\mathcal{T},\phi} = \int_0^t \left[(\mathbf{0}_{d-1})^\top \quad \mathbb{I}_{X_u \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}} \right] d\hat{W}_u$$

for all $t \in [0, S]$. The left-hand side follows from Definition 2.2.2: $\{D_t V_t^{\mathcal{T},\phi}\}$ is a $(\mathbb{Q}, \{\mathcal{F}_t\})$ -martingale. The initial capital of the portfolio is clearly zero; using discounted portfolio dynamics (Equation 2.8), we have

$$\begin{aligned} \int_0^t D_u \phi_u^\top \nabla P_u^\top \sigma(X_u) dW_u &= \int_0^t \left[(\mathbf{0}_{d-1})^\top \quad \mathbb{I}_{X_u \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}} \right] d\hat{W}_u \\ \Rightarrow \int_0^t D_u \phi_u^\top \nabla P_u^\top \hat{\sigma}(X_u) d\hat{W}_u &= \int_0^t \left[(\mathbf{0}_{d-1})^\top \quad \mathbb{I}_{X_u \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}} \right] d\hat{W}_u \end{aligned}$$

for all $t \in [0, S]$. It follows straightforwardly from Lemma 3.1.4 that the final entry of the left-hand diffusion process $[D_u \phi_u^\top \nabla P_u^\top \hat{\sigma}(X_u)]_d$ is zero when $X_u \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}$. This is precisely when the final entry of the right-hand diffusion process is equal to one. But because the integrals are equal for all $t \in [0, S]$, the two diffusion processes must be $\lambda \otimes \mathbb{P}$ -almost surely equal (see Filipović [2009, Ch.4.1] — we refer to this *identification result* below). The set of values of the state process that cause them to differ is therefore $X_*(\lambda \otimes \mathbb{P})$ -null. This implies that $\bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}$ is $X_*(\lambda \otimes \mathbb{P})$ -null. By Definition 3.1.7, the TSM contains no unspanned factors, completing the proof of Theorem 3.1.9.

3.3.2 Proof of Theorem 3.2.6

Intuitively, a USV factor means that there is stochastic variation affecting ZCB volatility but not ZCB prices. Many \mathcal{F}_S^P -measurable claims — such as bond options, caps and swaptions — are volatility-sensitive, and thus ZCBs are inadequate to fully hedge (and therefore attain) them if USV factors are present. The approach of the proof is to consider the bond covariance function (Definition 3.2.1) as a claim — like the claim in Equation 3.34 above, the bond-covariance claim will be shown to depend on the unspanned variation in an explicit way, and thus be unattainable.

We start the proof by confirming that the bond-covariance T^* -claim $v(\tau_1, \tau_2, X_{T^*})$ is $\mathcal{F}_{T^*}^P$ -measurable, making its attainability necessary for economic completeness (assuming the bounded-

²²Measurability is clear: the integrand, a function of the state process, is $\{\mathcal{F}_t\}$ -adapted, making the integral likewise adapted (and the discount factor, which depends on the state process, is \mathcal{F}_S -measurable). The boundedness follows from observing that $\Theta^* D_S = \int_{\mathcal{A}} d\hat{W}_u^{(d)}$, where $\mathcal{A} = \{t \in [0, S] | X_t \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}\}$. Therefore, $\Theta^* D_S$ is a normal random variable with zero mean and a variance of $\lambda(\mathcal{A})$. Thus, $|\Theta^* D_S|$ follows the *folded normal distribution*, and has a (finite) expectation of $\sqrt{\frac{2\lambda(\mathcal{A})}{\pi}}$ (see Leone, Nelson & Nottingham [1961]).

ness condition Equation 2.11 is met — see Assumption 3.3.4 below). We then show — given the assumption of a USV factor and the absence of partially spanned factors — that this claim cannot be attained.

The T^* -claim $v(\tau_1, \tau_2, X_{T^*})$ is clearly \mathcal{F}_{T^*} -measurable. It is also $\mathcal{F}_{T^*}^P$ -measurable, because, as mentioned in Section 3.2, $v(\tau_1, \tau_2, X_{T^*})$ is the time- T^* rate at which the two processes $\{P_{t, T^*+\tau_1}\}$ and $\{P_{t, T^*+\tau_2}\}$ are accumulating quadratic covariation. More specifically, quadratic covariation at time t , denoted

$$[P_{\cdot, T^*+\tau_1}, P_{\cdot, T^*+\tau_2}]_t,$$

is $\mathcal{F}_{T^*}^P$ -measurable for all $t \in [0, T^*]$, because it is a function of the ZCB price processes over $[0, T^*]$. Note that the quadratic covariation process $\{[P_{\cdot, T^*+\tau_1}, P_{\cdot, T^*+\tau_2}]_t\}$ is differentiable with respect to t ; the process is given by²³

$$[P_{\cdot, T^*+\tau_1}, P_{\cdot, T^*+\tau_2}]_t = \int_0^t \nabla P(T^* + \tau_1 - u, X_u)^\top \sigma(X_u) \sigma(X_u)^\top \nabla P(T^* + \tau_2 - u, X_u) du,$$

so that

$$\begin{aligned} \frac{\partial}{\partial t} [P_{\cdot, T^*+\tau_1}, P_{\cdot, T^*+\tau_2}]_t &= \nabla P(T^* + \tau_1 - t, X_t)^\top \sigma(X_t) \sigma(X_t)^\top \nabla P(T^* + \tau_2 - t, X_t) \\ &= v(T^* + \tau_1 - t, T^* + \tau_2 - t, X_{T^*}), \end{aligned}$$

where the final line uses Definition 3.2.1. Because the quadratic covariation process is $\mathcal{F}_{T^*}^P$ -measurable for all $t \in [0, T^*]$, its time- t rate of change (i.e., its partial derivative with respect to t , evaluated at any $t \in [0, T^*]$) is therefore also $\mathcal{F}_{T^*}^P$ -measurable. Evaluating the derivative at $t = T^*$ gives $v(\tau_1, \tau_2, X_{T^*})$, which is therefore $\mathcal{F}_{T^*}^P$ -measurable.

Then, from Definition 3.2.5, the existence of at least one USV factor implies that $\mathcal{X}_{\mathcal{V}, \mathcal{U}}^{(i)}$ is $X_*(\lambda \otimes \mathbb{P})$ -non-null for some $i > 0$. Therefore $\bigcup_{j=1}^d \mathcal{X}_{\mathcal{V}, \mathcal{U}}^{(j)}$ is $X_*(\lambda \otimes \mathbb{P})$ -non-null. Let \mathcal{X}^* equal $\bigcup_{j=1}^d \mathcal{X}_{\mathcal{V}, \mathcal{U}}^{(j)}$, minus any points where $\epsilon(\cdot)$ is not left-continuous (recall that $\epsilon(\cdot)$ is used to orthonormally transform the factor process in Lemma 3.1.4 and that, by Assumption 3.1.6, any points lacking left-continuity are $X_*(\lambda \otimes \mathbb{P})$ -null). Using Tonelli's theorem, we have

$$X_*(\lambda \otimes \mathbb{P})(\mathcal{X}^*) = \int_0^S \mathbb{P}(X_t \in \mathcal{X}^*) \lambda(dt) > 0,$$

which implies that there is a λ -non-null (and therefore non-empty) subset $\mathbb{T} \subseteq [0, S]$ such that

$$\mathbb{P}(X_t \in \mathcal{X}^*) > 0 \tag{3.35}$$

for all $t \in \mathbb{T}$.

From Corollary 3.2.4, we have the following for $x \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{V}, \mathcal{U}}^{(j)}$ and therefore for $x \in \mathcal{X}^*$:

$$[\nabla v(\tau_1^*, \tau_2^*, x)^\top \hat{\sigma}(x)]_d \neq 0 \tag{3.36}$$

for some $\tau_1^*, \tau_2^* > 0$. Now let T^* be any element of \mathbb{T} and, for all $t \in [0, T^*]$, define

$$q(t, X_t) = \mathbb{E}^{\mathbb{Q}} \left[\frac{D_{T^*}}{D_t} v(\tau_1^*, \tau_2^*, X_{T^*}) \middle| \mathcal{F}_t \right],$$

²³See Crépey [2013, Ch.3.2] for the well-known formula relating quadratic covariation to the processes' dynamics; see Equation 2.6 for the relevant dynamics.

noting that the Markov property of the state process under \mathbb{Q} (Assumption 2.1.5) ensures that the conditional expectation is indeed given by a function of X_t . Note also the terminal condition on $q(\cdot, \cdot)$: because $v(\tau_1^*, \tau_2^*, X_{T^*})$ is \mathcal{F}_{T^*} -measurable, we have that $q(T^*, x) = v(\tau_1^*, \tau_2^*, x)$, for all $x \in \mathcal{X}$. From this terminal condition and from Equation 3.36, we have that

$$[\nabla q(T^*, x)^\top \hat{\sigma}(x)]_d \neq 0$$

for $x \in \mathcal{X}^*$. Because T^* was taken from \mathbb{T} , from Equation 3.35 we have that

$$\mathbb{P}\left([\nabla q(T^*, X_{T^*})^\top \hat{\sigma}(X_{T^*})]_d \neq 0 \text{ and } \epsilon(X_{T^*}) \text{ is locally left-continuous}\right) > 0. \quad (3.37)$$

Assumption 3.3.4. $\mathbb{E}^\mathbb{Q}[|D_{T^*} v(\tau_1^*, \tau_2^*, X_{T^*})|] < \infty$, so that $q(\cdot, \cdot)$ is well-defined. Furthermore, $q(\cdot, \cdot)$ is a $\mathcal{C}^{1,2}$ function.

$q(\cdot, \cdot)$ is well known to be the Feynman-Kac stochastic representation of the solution to a corresponding parabolic partial differential equation (PDE). Under mild conditions given by Karatzas & Shreve [2012, Ch.4.4], a $\mathcal{C}^{1,2}$ solution to this equation exists. These conditions include some boundedness and continuity constraints on the short rate process, which are implicit in Assumption 3.3.2, and a growth condition on $v(\tau_1^*, \tau_2^*, \cdot)$.

Given Assumption 3.3.4, Itô's Lemma can be applied to $\{q(t, X_t)\}$, resulting in

$$q(t, X_t) = q(0, X_0) + \int_0^t r_u q(u, X_u) du + \int_0^t \nabla q(u, X_u)^\top \hat{\sigma}(X_u) d\hat{W}_u$$

for all $t \in [0, T^*]$, where the above-mentioned PDE is used to replace the numerous terms in the drift process. Now consider a general admissible portfolio process $\{V_t^{\mathcal{T}, \phi}\}$. From Equation 2.9, we have

$$V_t^{\mathcal{T}, \phi} = x + \int_0^t r_u V_u^{\mathcal{T}, \phi} du + \int_0^t \phi_u^\top \nabla P_u^{\mathcal{T}} \hat{\sigma}(X_u) d\hat{W}_u$$

for all $t \in [0, T^*]$. Note that if $\{V_t^{\mathcal{T}, \phi}\}$ were to attain $v(\tau_1^*, \tau_2^*, X_{T^*})$, it would follow that $V_t^{\mathcal{T}, \phi} = q(t, X_t)$ for all $t \in [0, T^*]$ (from the definition of $q(\cdot, \cdot)$ above and the martingale property of admissible discounted portfolio processes in Definition 2.2.2). We claim that the above two diffusion processes $\{\nabla q(t, X_t)^\top \hat{\sigma}(X_t)\}$ and $\{\phi_t^\top \nabla P_t^{\mathcal{T}} \hat{\sigma}(X_t)\}$ *cannot* be $\lambda \otimes \mathbb{P}$ -almost surely equal. Using the contrapositive of the identification result used in Section 3.3.1 above, it would follow that $V_t^{\mathcal{T}, \phi}$ is not (almost-surely) equal to $q(t, X_t)$ for all $t \in [0, T^*]$, and then that $v(\tau_1^*, \tau_2^*, X_{T^*})$ is not attainable, causing the TSM to be economically incomplete.

Note that $\nabla q(t, x)^\top \hat{\sigma}(x)$ is left-continuous wherever $\epsilon(x)$ is left-continuous (this follows from Assumptions 2.1.5 and 3.3.4, and the definition of $\hat{\sigma}(x) = \sigma(x)\epsilon(x)$ in Lemma 3.1.4). It then follows from Equation 3.37 that there is a non-zero probability that $[\nabla q(t, X_t)^\top \hat{\sigma}(X_t)]_d$ is non-zero for all $t \in [T^* - \lambda^*, T^*]$ (where λ^* is some positive constant). Therefore, $\{[\nabla q(t, X_t)^\top \hat{\sigma}(X_t)]_d\}$ is *not* $\lambda \otimes \mathbb{P}$ -almost everywhere equal to zero.

The proof of Theorem 3.2.6 is then completed by comparing this to $\{[\phi_t^\top \nabla P_t^{\mathcal{T}} \hat{\sigma}(X_t)]_d\}$. Note that, from the theorem's assumption of the absence of partially spanned factors, there must exist at least one fully unspanned factor (the existence of a USV factor implies the existence of an unspanned factor; it is fully unspanned by assumption). Lemma 3.1.4 then ensures that $\{[\nabla P(\tau, X_t) \hat{\sigma}(X_t)]_d\}$, and therefore $\{[\phi_t^\top \nabla P_t^{\mathcal{T}} \hat{\sigma}(X_t)]_d\}$, is $\lambda \otimes \mathbb{P}$ -almost everywhere zero.²⁴

²⁴Corollary 3.2.4 gives a particular way that the basis vectors are selected in the proof of Lemma 3.1.4 — the statement given by Lemma 3.1.4 is not affected by this.

3.3.3 Proof of Theorem 3.2.10

The theorem assumes that an \mathcal{S} -transformation is possible and that partially spanned state variables are absent. We can therefore follow Equation 3.18 in partitioning the transformed state process $\{\hat{X}_t\}$ into spanned and unspanned components $\{Y_t\}$ and $\{U_t\}$, and then Lemma 3.1.17, by giving ZCB prices in terms of a reduced bond pricing function:

$$P_{tT} = \tilde{P}(T - t, Y_t).$$

Recall that p and \bar{p} give the number of unspanned factors and unspanned state variables respectively. Using the theorem's assumption that partially spanned factors are absent, we can apply Lemma 3.1.18, which gives the dynamics of the spanned component as

$$\begin{aligned} dY_t &= \mu^Y(\hat{X}_t) dt + \begin{bmatrix} \sigma^Y(\hat{X}_t) & \mathbf{0}_{(n-\bar{p}) \times p} \end{bmatrix} d\hat{W}_t \\ &= \mu^Y(\hat{X}_t) dt + \sigma^Y(\hat{X}_t) dW_t^Y, \end{aligned} \quad (3.38)$$

where $\{W_t^Y\}$ denotes the first $d - p$ scalar components of (d -dimensional) $\{\hat{W}_t\}$. It is convenient to let $m = d - p$ and $\bar{m} = n - \bar{p}$, so that $\{W_t^Y\}$ is a m -dimensional and $\{Y_t\}$ is \bar{m} -dimensional.

The proof is organised into four subsections. The central idea of the proof is that $\{Y_t\}$ can be shown to be adapted to a filtration generated by an m -dimensional \mathbb{Q} -Brownian motion. It useful to give an overview of each subsection, and to note how they relate to this central idea:

- *Lower-triangular orthonormal transformation:* $\{Y_t\}$ is not necessarily adapted to the filtration generated by $\{W_t^Y\}$. The first step to is construct a further orthonormal transformation (in addition to the one performed in Lemma 3.1.4) that defines $\{\hat{W}_t^Y\}$, which is suitable for the central idea of the proof.
- *Invertible matrix processes:* two almost-surely-invertible bond-sensitivity matrix processes are needed for the proof. One is used for the eventual hedging portfolio (like the one given in Section 3.3.1 in the proof of Theorem 3.1.9); one is used for a different purpose in the following subsection.
- *Adaptedness of the spanned component:* here we show that $\{Y_t\}$ is adapted to the filtration generated by $\{\hat{W}_t^Y\}$. We do this by showing that $\{Y_t\}$ satisfies a stochastic differential equation (SDE) where $\{\hat{W}_t^Y\}$ is the stochastic integrator. This requires that the drift and diffusion processes can be represented without the unspanned component $\{U_t\}$. To show this, one of the invertible bond-sensitivity matrix processes from the second subsection is required. For the diffusion, the theorem's assumption of a lack of USV state variables is instrumental.
- *Attainability of a general claim:* with the adaptedness of the spanned component established, the martingale representation theorem can be applied to show that an *economic* claim (claims based on the term structure — recall the notion of economic completeness from Definition 2.2.6) can be attained, despite the presence of any unspanned factors (which cause the model to be mathematically incomplete, as per Theorem 3.1.9).

Lower-triangular Orthonormal Transformation

The spanned component's diffusion process $\{\sigma^Y(\hat{X}_t)\}$ is an $\bar{m} \times m$ matrix process. Suppose, for an initial example, that $\bar{m} = m$ and that $\{\sigma^Y(\hat{X}_t)\}$ is invertible. Then one can define a matrix

process $\{C_t\}$ such that C_t is the *Cholesky decomposition* of the variance matrix $\sigma^Y(\hat{X}_t)\sigma^Y(\hat{X}_t)^\top$.²⁵ Then consider defining

$$\hat{W}_t^Y = \int_0^t C_u^{-1} \sigma^Y(\hat{X}_u) dW_u^Y,$$

which is easily seen to be a Brownian motion (by using Lévy's characterisation like with the orthonormal transformations below Definition 2.2.8). One could then modify the dynamics in Equation 3.38:

$$\begin{aligned} dY_t &= \mu^Y(\hat{X}_t) dt + C_t C_t^{-1} \sigma^Y(\hat{X}_t) dW_t^Y \\ &= \mu^Y(\hat{X}_t) dt + C_t d\hat{W}_t^Y. \end{aligned}$$

The motivation behind this is that the modified diffusion process is lower triangular (see Footnote 25), which is needed in the third subsection of the proof (with lower triangularity, we can show that $\{Y_t\}$ satisfies an SDE and is therefore adapted).

In the general case, we have $\bar{m} \geq m$ and the assurance that $\sigma^Y(\hat{X}_t)$ is full rank (see Lemma 3.1.18). We now describe a procedure to construct an orthonormal transformation (Definition 2.2.8) that results in the modified diffusion matrix being *generalised lower triangular*, in a sense we clarify below.

Note firstly that the \bar{m} rows of $\sigma^Y(\hat{X}_t)$ span \mathbb{R}^m . Counting from the first row downward, m of the \bar{m} rows will contribute an extra dimension to the m -dimensional row space of $\sigma^Y(\hat{X}_t)$ (as it is expanded down the rows) — we refer to these as the *rank-adding* rows.

Now define a matrix-valued function $\bar{\epsilon} : \hat{\mathcal{X}} \rightarrow \mathbb{R}^{m \times m}$, beginning with the final column, which, for each $\hat{x} \in \hat{\mathcal{X}}$, must be set to be of unit length and orthogonal to the first $(m-1)$ rank-adding rows of $\sigma^Y(\hat{x})$. The first $(m-1)$ rows of $\sigma^Y(\hat{x})$ might be linearly independent and span an $(m-1)$ -dimensional linear subspace of \mathbb{R}^m — in this case, the final column of $\bar{\epsilon}(\hat{x})$ must be one of the two unit vectors in the one-dimensional orthogonal complement of this subspace. If the first $(m-1)$ rows of $\sigma^Y(\hat{x})$ contain a linear dependence, further rows must be considered until an $(m-1)$ -dimensional subspace is spanned (giving the first $(m-1)$ rank-adding rows) — this gives a one-dimensional orthogonal complement in which the final column of $\bar{\epsilon}(\hat{x})$ is selected.

The second last column of $\bar{\epsilon}(\hat{x})$ is chosen in the same way, except that the first rank-adding row is ignored, and orthogonality to the already-determined final column replaces this constraint. That is, an $(m-1)$ -dimensional linear subspace is determined as the span of final column of $\bar{\epsilon}(\hat{x})$ as well as the second, third, ..., $(m-2)$ th and $(m-1)$ th rank-adding rows of $\sigma^Y(\hat{x})$, and then the second last column of $\bar{\epsilon}(\hat{x})$ is set as a unit-length element of the orthogonal complement of this subspace.²⁶ This process is continued: the i th column of $\bar{\epsilon}(\hat{x})$ from the right-hand side, is set to be a unit vector orthogonal to the first i th, $(i+1)$ th, ..., $(m-1)$ th rank-adding rows of $\sigma^Y(\hat{x})$ and also orthogonal to the $i-1$ rightmost columns of $\bar{\epsilon}(\hat{x})$.

This construction has the consequence that the first i rank-adding rows of $\sigma^Y(\hat{x})$ are orthogonal to the $(i+1)$ th, $(i+2)$ th, ..., $(m-1)$ th and m th columns of $\bar{\epsilon}(\hat{x})$. It also ensures that $\bar{\epsilon}(\hat{x})$ is an orthogonal matrix for each $\hat{x} \in \hat{\mathcal{X}}$, and is therefore suitable to transform $\{W_t^Y\}$ as per Definition

²⁵The Cholesky decomposition — a lower-triangular matrix C_t such that $C_t C_t^\top = \sigma^Y(\hat{X}_t)\sigma^Y(\hat{X}_t)^\top$ — would exist (uniquely) in this case, and also be invertible. See Golub & Van Loan [2012, Ch.4.2].

²⁶Because the final column of $\bar{\epsilon}(\hat{x})$ was set to be orthogonal to these rank-adding rows, an $(m-1)$ -dimensional linear subspace is necessarily obtained.

2.2.8:

$$\hat{W}_t^Y = \int_0^t \bar{\epsilon}(\hat{X}_u)^{-1} dW_u^Y. \quad (3.39)$$

This allows us to express the dynamics of the spanned component in terms of $\{\hat{W}_t^Y\}$:

$$\begin{aligned} dY_t &= \mu^Y(\hat{X}_t) dt + \sigma^Y(\hat{X}_t) \bar{\epsilon}(\hat{X}_t) \bar{\epsilon}(\hat{X}_t)^{-1} dW_t^Y \\ &= \mu^Y(\hat{X}_t) dt + \sigma^Y(\hat{X}_t) \bar{\epsilon}(\hat{X}_t) d\hat{W}_t^Y \\ &= \mu^Y(\hat{X}_t) dt + \hat{\sigma}^Y(\hat{X}_t) d\hat{W}_t^Y, \end{aligned} \quad (3.40)$$

where we have defined $\hat{\sigma}^Y(\hat{X}_t) = \sigma^Y(\hat{X}_t) \bar{\epsilon}(\hat{X}_t)$. The row of $\hat{\sigma}^Y(\hat{X}_t)$ that corresponds to the first rank-adding row of $\sigma^Y(\hat{X}_t)$ has only one non-zero entry, in the first column.^{27,28} The row of $\hat{\sigma}^Y(\hat{X}_t)$ that corresponds to the second rank-adding row of $\sigma^Y(\hat{X}_t)$ has two non-zero entries, in the first two columns (the second of which is necessarily non-zero), and so on.²⁹

This is the generalised lower triangular form mentioned above. If $\bar{m} = m$ (i.e., if $\sigma^Y(\hat{X}_t)$ and $\hat{\sigma}^Y(\hat{X}_t)$ are square) the construction of $\bar{\epsilon}(\cdot)$ ensures that

$$[\hat{\sigma}^Y(\hat{X}_t)]_{ij} = 0 \quad (3.41)$$

for $i < j$ (that is, $\hat{\sigma}^Y(\hat{X}_t)$ is lower triangular). This is because, in this case, the rank-adding rows are simply the rows, so the first i rows of $\sigma^Y(\hat{X}_t)$ are orthogonal to the $(i+1)$ th, $(i+2)$ th, ..., $(m-1)$ th and m th columns of $\bar{\epsilon}(\hat{X}_t)$. For example, suppose the spanned component satisfies

$$dY_t = d \begin{bmatrix} \hat{X}_t^{(1)} \\ \hat{X}_t^{(2)} \end{bmatrix} = \begin{bmatrix} \sqrt{\hat{X}_t^{(1)}} & \sigma \\ \sqrt{\hat{X}_t^{(2)}} & \sqrt{\hat{X}_t^{(3)}} \end{bmatrix} dW_t^Y$$

for some non-zero constant σ and positive transformed state process $\{\hat{X}_t\}$. Following the above, we can define

$$\bar{\epsilon}(\hat{X}_t) = \begin{bmatrix} \frac{\sqrt{\hat{X}_t^{(1)}}}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} & \frac{-\sigma}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} \\ \frac{\sigma}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} & \frac{\sqrt{\hat{X}_t^{(1)}}}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} \end{bmatrix},$$

which is orthogonal and is used to define $\{\hat{W}_t^Y\}$ with Equation 3.39, so that

$$\begin{aligned} dY_t &= \begin{bmatrix} \sqrt{\hat{X}_t^{(1)}} & \sigma \\ \sqrt{\hat{X}_t^{(2)}} & \sqrt{\hat{X}_t^{(3)}} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{\hat{X}_t^{(1)}}}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} & \frac{-\sigma}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} \\ \frac{\sigma}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} & \frac{\sqrt{\hat{X}_t^{(1)}}}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} \end{bmatrix} d\hat{W}_t^Y \\ &= \begin{bmatrix} \sqrt{\hat{X}_t^{(1)} + \sigma^2} & 0 \\ \frac{\sqrt{\hat{X}_t^{(1)}} \sqrt{\hat{X}_t^{(2)} + \sigma \sqrt{\hat{X}_t^{(3)}}}}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} & \frac{\sqrt{\hat{X}_t^{(1)}} \sqrt{\hat{X}_t^{(3)}} - \sigma \sqrt{\hat{X}_t^{(2)}}}{\sqrt{\hat{X}_t^{(1)} + \sigma^2}} \end{bmatrix} d\hat{W}_t^Y, \end{aligned}$$

²⁷This is because the first rank-adding row of $\sigma^Y(\hat{X}_t)$ is orthogonal to all columns in $\bar{\epsilon}(\hat{X}_t)$ except the first.

²⁸Note also that this entry is necessarily non-zero — if it were not, the relevant row of $\hat{\sigma}^Y(\hat{X}_t)$ would have to comprised of zeros, precluding it from adding rank.

²⁹The i th entry of the row corresponding to the i th rank-adding row of $\sigma^Y(\hat{x})$ must be non-zero $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere — if it were not, the i th rank-adding row of $\sigma^Y(\hat{x})$ would, like the earlier rank-adding rows, be orthogonal to the i th column of $\bar{\epsilon}(\hat{X}_t)$ and would not contribute rank. We have, from Lemma 3.1.18, that $\sigma^Y(\hat{x})$ is full rank $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere; excluding any points where it is not full rank, this must hold for $i = 1, 2, \dots, m$.

which gives the desired lower triangularity.

For the cases where $\bar{m} > m$, it is useful to consider some static examples — evaluating at some \hat{x} , the following is an example of the generalised-lower-triangular transformation:

$$\sigma^Y(\hat{x})\bar{\epsilon}(\hat{x}) = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{7}{\sqrt{5}} & \frac{1}{\sqrt{5}} \\ 1 & 0 \end{bmatrix} = \hat{\sigma}^Y(\hat{x}).$$

Here the first two rows are the rank-adding rows. The corresponding rows of $\hat{\sigma}^Y(\hat{x})$ together form a lower-triangular matrix. At some other \hat{x} , the second row might not contribute rank:

$$\sigma^Y(\hat{x})\bar{\epsilon}(\hat{x}) = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{10}} & \frac{-3}{\sqrt{10}} \\ \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ \frac{7}{\sqrt{10}} & \frac{-1}{\sqrt{10}} \end{bmatrix} = \hat{\sigma}^Y(\hat{x}).$$

Finally, suppose that the first row is not a rank-adding row:

$$\sigma^Y(\hat{x})\bar{\epsilon}(\hat{x}) = \begin{bmatrix} 0 & 0 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{5}} & \frac{-2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ \frac{7}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{bmatrix} = \hat{\sigma}^Y(\hat{x}).$$

Note that it is not necessary to implement this transformation in practice — we simply need its existence for this proof. Nevertheless, it does have the following implication.

Remark 3.3.5. When specifying a diffusion matrix, one may as well specify it in a (generalised) lower-triangular form — the above transformation demonstrates that no generality is gained from other specifications, because, after a redefinition of the Brownian motions, the model can simply be recast with a lower-triangular diffusion matrix. This fact is used in Chapter 5.

Invertible Matrix Processes

Two invertible matrix processes are required before $\{Y_t\}$ is shown to be adapted to $\{\hat{W}_t^Y\}$. It is useful to first define \mathcal{U} - and \mathcal{S} -partitions of the transformed state space $\hat{\mathcal{X}}$ — for $i = 0, 1, \dots, d$, define

$$\hat{\mathcal{X}}_{\mathcal{U}}^{(i)} = \{\hat{x} \in \hat{\mathcal{X}} \mid B(\hat{x}) \in \mathcal{X}_{\mathcal{U}}^{(i)}\},$$

so that $X_t \in \mathcal{X}_{\mathcal{U}}^{(i)}$ if and only if $\hat{X}_t \in \hat{\mathcal{X}}_{\mathcal{U}}^{(i)}$. In the same spirit, define, for $i = 0, 1, \dots, n$,

$$\hat{\mathcal{X}}_{\mathcal{S}}^{(i)} = \{\hat{x} \in \hat{\mathcal{X}} \mid B(\hat{x}) \in \mathcal{X}_{\mathcal{S}}^{(i)}\}.$$

The $d \times n$ matrix process $\{\nabla P_t^T\}$ defined in Section 2.2 to specialised to a time-homogeneous context in Section 3.3.1. Here we specialise further to make use of the reduced bond pricing function, and also to accommodate two maturity processes $\{\mathcal{T}_t\}$ and $\{\bar{\mathcal{T}}_t\}$, which are m - and \bar{m} -dimensional respectively. That is, we define an $m \times \bar{m}$ -matrix process $\{\nabla \tilde{P}_t^T\}$ with entries

$$[\nabla \tilde{P}_t^T]_{ij} = \frac{\partial \tilde{P}([T_t]_i - t, y)}{\partial y^{(j)}} \Big|_{y=Y_t},$$

as well as an $\bar{m} \times \bar{m}$ -matrix process $\{\nabla \tilde{P}_t^{\bar{\mathcal{T}}}\}$ with entries

$$[\nabla \tilde{P}_t^{\bar{\mathcal{T}}}]_{ij} = \frac{\partial \tilde{P}([\bar{\mathcal{T}}_t]_i - t, y)}{\partial y^{(j)}} \Big|_{y=Y_t}.$$

We wish to show that there exist maturity processes $\{\mathcal{T}_t\}$ and $\{\bar{\mathcal{T}}_t\}$ ensuring that $\{\nabla \tilde{P}_t^{\mathcal{T}} \sigma^Y(\hat{X}_t)\}$ and $\{\nabla \tilde{P}_t^{\bar{\mathcal{T}}}\}$ are $\lambda \otimes \mathbb{P}$ -almost surely invertible. This requires two analogues of Lemma 3.3.1, and then an analogue of Assumption 3.3.2.

Corollary 3.3.6. Assuming $\hat{X}_t \in \bigcup_{j=1}^p \hat{\mathcal{X}}_{\mathcal{U}}^{(j)}$, there exists an m -dimensional maturity process $\{\mathcal{T}_t\}$ such that $\nabla \tilde{P}_t^{\mathcal{T}} \sigma^Y(\hat{X}_t)$ is invertible.

Corollary 3.3.7. Assuming $\hat{X}_t \in \bigcup_{j=0}^{\bar{p}} \hat{\mathcal{X}}_{\mathcal{S}}^{(j)}$, there exists an \bar{m} -dimensional maturity process $\{\bar{\mathcal{T}}_t\}$ such that $\nabla \tilde{P}_t^{\bar{\mathcal{T}}}$ is invertible.

These follow from reasoning parallel to that in the proof of Lemma 3.3.1. If $\hat{X}_t \in \bigcup_{j=0}^p \hat{\mathcal{X}}_{\mathcal{U}}^{(j)}$, then the m maturities \mathcal{T}_t must exist such that $\nabla \tilde{P}_t^{\mathcal{T}} \sigma^Y(\hat{X}_t)$ is invertible, because if they did not, it would imply that \hat{X}_t is in fact an element of $\bigcup_{j=p+1}^d \hat{\mathcal{X}}_{\mathcal{U}}^{(j)}$.³⁰ Similarly, an inability to find \bar{m} maturities $\bar{\mathcal{T}}_t$ ensuring the invertibility of $\nabla \tilde{P}_t^{\bar{\mathcal{T}}}$ indicates that $\hat{X}_t \in \bigcup_{j=\bar{p}+1}^n \hat{\mathcal{X}}_{\mathcal{S}}^{(j)}$.

Corollaries 3.3.6 and 3.3.7 show that suitable maturities can be found at any given time t . Section 3.3.1 discusses how this may be insufficient to guarantee suitable maturities processes (which cannot select maturities for each time t , but must specify maturities in advance for intervals of time points). However, stronger results than Corollaries 3.3.6 and 3.3.7 hold in practice — not only do suitable maturities exist, one finds that *any set of distinct maturities* is suitable. This *general maturity availability*, in the context of this section, is formalised in the following.

Assumption 3.3.8. If $\hat{X}_t \in \bigcup_{j=0}^p \hat{\mathcal{X}}_{\mathcal{U}}^{(j)}$, any m -dimensional maturity process $\{\mathcal{T}_t\}$, provided that the m maturities \mathcal{T}_t are distinct, gives rise to an invertible matrix $\nabla \tilde{P}_t^{\mathcal{T}} \sigma^Y(\hat{X}_t)$. Similarly, if $\hat{X}_t \in \bigcup_{j=0}^{\bar{p}} \hat{\mathcal{X}}_{\mathcal{S}}^{(j)}$, any \bar{m} -dimensional maturity process $\{\bar{\mathcal{T}}_t\}$, provided that the \bar{m} maturities $\bar{\mathcal{T}}_t$ are distinct, gives rise to an invertible matrix $\nabla \tilde{P}_t^{\bar{\mathcal{T}}}$.

From the same reasoning in Section 3.3.1, this is sufficient to ensure that $\lambda \otimes \mathbb{P}$ -almost surely invertible $\{\nabla \tilde{P}_t^{\mathcal{T}} \sigma^Y(\hat{X}_t)\}$ and $\{\nabla \tilde{P}_t^{\bar{\mathcal{T}}}\}$, based on maturity processes $\{\mathcal{T}_t\}$ and $\{\bar{\mathcal{T}}_t\}$, can be attained. Appendix B discusses relaxing Assumption 3.3.8.

Remark 3.3.9. Assumption 3.3.8 (which is an exact analogue of Assumption 3.3.2, accounting for the fully unspanned factors and state variables potentially present in Section 3.3.3) holds in practice because bond pricing functions are not restrictive in their availability of suitable maturities.

Consider Example 3, where bonds are given in terms of a single state variable via Equation 3.17. Using the reduced bond pricing function, we have

$$\nabla \tilde{P}(\tau, y) = \frac{\partial \tilde{P}(\tau, y)}{\partial y} = -2b(\tau)y\tilde{P}(\tau, y),$$

³⁰In the proof of Lemma 3.3.1, this was shown by considering, in effect, the intersection of the kernels of $\nabla \tilde{P}_t^{\mathcal{T}} \sigma^Y(\hat{X}_t)$ for all possible maturity choices. Similar reasoning applies here, except that any non-zero element of the kernel intersection δ is not itself an element of $\mathcal{U}^{B(\hat{X}_t)}$, but does correspond to such an element, namely

$$\delta^* = \epsilon(B(\hat{X}_t))^\top \begin{bmatrix} \bar{\epsilon}(\hat{X}_t)^\top \delta \\ \mathbf{0}_p \end{bmatrix}.$$

Furthermore, one can see that the membership of δ^* implies that $\mathcal{U}^{B(\hat{X}_t)}$ has dimension larger than p , because it follows from the construction of $\epsilon(\cdot)$ (in Lemma 3.1.4) that for any $x \in \mathbb{R}^p$,

$$\epsilon(B(\hat{X}_t))^\top \begin{bmatrix} \mathbf{0}_m \\ x \end{bmatrix} \in \mathcal{U}^{B(\hat{X}_t)}.$$

where $b(\tau) \neq 0$ for $\tau > 0$. This is equal to zero only when $y = 0$, which corresponds precisely to $\hat{X}_t \in \hat{\mathcal{X}}_{\mathcal{U}}^{(2)} = \bigcup_{j=p+1}^d \hat{\mathcal{X}}_{\mathcal{U}}^{(j)}$ and to $\hat{X}_t \in \hat{\mathcal{X}}_{\mathcal{S}}^{(2)} = \bigcup_{j=\bar{p}+1}^n \hat{\mathcal{X}}_{\mathcal{S}}^{(j)}$. For all other states, *any maturity choice* gives rise to an invertible (i.e., non-zero) $\nabla \tilde{P}_t^{\mathcal{T}} \sigma^Y(\hat{X}_t)$ and $\nabla \tilde{P}_t^{\mathcal{T}}$, so that Assumption 3.3.8 is satisfied. One can then select any maturity for the first value of the piecewise-constant $\{\mathcal{T}_t\}$ and $\{\bar{\mathcal{T}}_t\}$ (and then reset it when or before the corresponding ZCB expires). Because the event $Y_t = 0$ (or, equivalently, $X_t = \mathbf{0}_2$) is $\lambda \otimes \mathbb{P}$ -null, the processes $\{\nabla \tilde{P}_t^{\mathcal{T}} \sigma^Y(\hat{X}_t)\}$ and $\{\nabla \tilde{P}_t^{\mathcal{T}}\}$ are $\lambda \otimes \mathbb{P}$ -almost surely invertible.

Adaptedness of the Spanned Component

We now aim to show that $\{Y_t\}$ can be written as the solution to an SDE based on the integrator $\{\hat{W}_t^Y\}$ (which will show that $\{Y_t\}$ is adapted to the filtration generated by $\{\hat{W}_t^Y\}$). We need to show that the drift and diffusion functions in Equation 3.40 depend only on $\{Y_t\}$ itself, and not on the unspanned component $\{U_t\}$. We do this in two lemmas, beginning with the diffusion function.

Lemma 3.3.10. $\hat{\sigma}^Y(\cdot)$ can be written as a function of the first \bar{m} components of its argument; that is, there exists a function $\tilde{\sigma}^Y(\cdot)$ such that

$$\hat{\sigma}^Y(Y_t) = \tilde{\sigma}^Y(\hat{X}_t).$$

Remark 3.3.11. The intuition of the lemma is straightforward: in the absence of USV state variables, the unspanned component U_t cannot have any influence on the spanned component's diffusion function. If they did have an effect, this would be inherited to the ZCBs themselves, causing the model to exhibit USV.

Proof. Define an $\bar{m} \times \bar{m}$ matrix process with

$$\hat{v}^{\mathcal{T}}(\hat{X}_t) = (\nabla \tilde{P}_t^{\mathcal{T}} \hat{\sigma}^Y(\hat{X}_t)) (\nabla \tilde{P}_t^{\mathcal{T}} \hat{\sigma}^Y(\hat{X}_t))^{\top}, \quad (3.42)$$

where the process $\{\nabla \tilde{P}_t^{\mathcal{T}}\}$ is from the second subsection of the proof. A typical element of $\hat{v}^{\mathcal{T}}(\hat{X}_t)$ is given by

$$[\hat{v}^{\mathcal{T}}(\hat{X}_t)]_{ij} = \nabla \tilde{P}([\bar{\mathcal{T}}_t]_i - t, Y_t)^{\top} \hat{\sigma}^Y(\hat{X}_t) \hat{\sigma}^Y(\hat{X}_t)^{\top} \nabla \tilde{P}([\bar{\mathcal{T}}_t]_j - t, Y_t), \quad (3.43)$$

which is the instantaneous covariance between ZCBs maturing at $[\bar{\mathcal{T}}_t]_i$ and $[\bar{\mathcal{T}}_t]_j$ (so that $\hat{v}^{\mathcal{T}}(\hat{X}_t)$ is an instantaneous covariance matrix of the bonds engaged in the maturity process $\{\bar{\mathcal{T}}_t\}$ from the second subsection). In particular, we claim that, for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all $\hat{x} \in \hat{\mathcal{X}}$,

$$[\hat{v}^{\mathcal{T}}(\hat{x})]_{ij} = v([\bar{\mathcal{T}}_t]_i - t, [\bar{\mathcal{T}}_t]_j - t, B(\hat{x})), \quad (3.44)$$

where $v(\cdot, \cdot, \cdot)$ is the bond covariance function defined in Equation 3.29. In light of Equation 3.29, this claim is intuitive, but confirming it formally requires us to relate the TSM to its pre-transformation form. We will discard a number of $\hat{X}_*(\lambda \otimes \mathbb{P})$ -null sets as we justify Equation 3.44.

For all $\hat{x} \in \hat{\mathcal{X}}$ except the $\hat{X}_*(\lambda \otimes \mathbb{P})$ -null points where $B(\cdot)$ is not differentiable or where $J_A(B(\cdot))$ does not exist, we have

$$\begin{aligned} \nabla P(\tau, B(\hat{x}))^{\top} \sigma(B(\hat{x})) &= \nabla P(\tau, B(\hat{x}))^{\top} \hat{\sigma}(B(\hat{x})) \epsilon(B(\hat{x}))^{\top} \\ &= \nabla \hat{P}(\tau, \hat{x})^{\top} J_A(B(\hat{x})) \hat{\sigma}(B(\hat{x})) \epsilon(B(\hat{x}))^{\top} \end{aligned}$$

from the definition of $\hat{\sigma}(\cdot)$ in Lemma 3.1.4 (recall that $\epsilon(B(\hat{x}))^{-1} = \epsilon(B(\hat{x}))^\top$) and from Equation 3.13, respectively. Let $y(\hat{x})$ denote the first \bar{m} elements of \hat{x} (so that $y(\hat{X}_t) = Y_t$), allowing us to continue by introducing the reduced bond pricing function:

$$\begin{aligned}\nabla P(\tau, B(\hat{x}))^\top \sigma(B(\hat{x})) &= \begin{bmatrix} \nabla \tilde{P}(\tau, y(\hat{x}))^\top & (\mathbf{0}_{\bar{p}})^\top \end{bmatrix} J_A(B(\hat{x})) \hat{\sigma}(B(\hat{x})) \epsilon(B(\hat{x}))^\top \\ &= \nabla \tilde{P}(\tau, y(\hat{x}))^\top \begin{bmatrix} \sigma^Y(\hat{x}) & \mathbf{0}_{\bar{m} \times p} \end{bmatrix} \epsilon(B(\hat{x}))^\top\end{aligned}$$

for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all \hat{x} . The first line follows from the $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere lack of sensitivity to the unspanned state variables (Definition 3.1.14), and the second from Equation 3.22 and the definition of $\sigma^Y(\cdot)$ in Lemma 3.1.18. Then, using the orthonormal transformation constructed above, we have

$$\nabla P(\tau, B(\hat{x}))^\top \sigma(B(\hat{x})) = \nabla \tilde{P}(\tau, y(\hat{x}))^\top \begin{bmatrix} \hat{\sigma}^Y(\hat{x}) \bar{\epsilon}(\hat{x})^\top & \mathbf{0}_{\bar{m} \times p} \end{bmatrix} \epsilon(B(\hat{x}))^\top,$$

which is used to express the original bond covariance function from Equation 3.29:

$$\begin{aligned}v(\tau_i, \tau_j, B(\hat{x})) &= \nabla P(\tau_i, B(\hat{x}))^\top \sigma(B(\hat{x})) \sigma(B(\hat{x}))^\top \nabla P(\tau_j, B(\hat{x})) \\ &= \nabla \tilde{P}(\tau_i, y(\hat{x}))^\top \begin{bmatrix} \hat{\sigma}^Y(\hat{x}) \bar{\epsilon}(\hat{x})^\top & \mathbf{0}_{\bar{m} \times p} \end{bmatrix} \epsilon(B(\hat{x}))^\top \epsilon(B(\hat{x})) \begin{bmatrix} \bar{\epsilon}(\hat{x}) \hat{\sigma}^Y(\hat{x})^\top \\ \mathbf{0}_{p \times (n-\bar{p})} \end{bmatrix} \nabla \tilde{P}(\tau_j, y(\hat{x})) \\ &= \nabla \tilde{P}(\tau_i, y(\hat{x}))^\top \begin{bmatrix} \hat{\sigma}^Y(\hat{x}) \bar{\epsilon}(\hat{x})^\top & \mathbf{0}_{\bar{m} \times p} \end{bmatrix} \begin{bmatrix} \bar{\epsilon}(\hat{x}) \hat{\sigma}^Y(\hat{x})^\top \\ \mathbf{0}_{p \times (n-\bar{p})} \end{bmatrix} \nabla \tilde{P}(\tau_j, y(\hat{x})) \\ &= \nabla \tilde{P}(\tau_i, y(\hat{x}))^\top \hat{\sigma}^Y(\hat{x}) \bar{\epsilon}(\hat{x})^\top \bar{\epsilon}(\hat{x}) \hat{\sigma}^Y(\hat{x})^\top \nabla \tilde{P}(\tau_j, y(\hat{x})) \\ &= \nabla \tilde{P}(\tau_i, y(\hat{x}))^\top \hat{\sigma}^Y(\hat{x}) \hat{\sigma}^Y(\hat{x})^\top \nabla \tilde{P}(\tau_j, y(\hat{x})).\end{aligned}$$

When evaluated at $\hat{x} = \hat{X}_t$, $\tau_i = [\tilde{T}_t]_i - t$ and $\tau_j = [\tilde{T}_t]_j = t$, this coincides exactly with Equation 3.43, confirming that Equation 3.44 holds.

For all $\hat{x} \in \hat{\mathcal{X}}$, define

$$\hat{v}(\tau_1, \tau_2, \hat{x}) = v(\tau_1, \tau_2, B(\hat{x})).$$

Then, using the chain rule, we have

$$\nabla \hat{v}(\tau_1, \tau_2, \hat{x})^\top = \nabla v(\tau_1, \tau_2, B(\hat{x}))^\top J_B(\hat{x}).$$

The final \bar{p} columns of $J_B(\hat{x})$ are $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere elements of $\mathcal{S}^{B(\hat{x})}$ (this follows from Equations 3.9 and 3.13, and is discussed further in the proof of Lemma 3.1.18). Because of the assumption of a lack of USV state variables, for any $\delta \in \mathcal{S}^{B(\hat{x})}$ and any $\tau_1, \tau_2 > 0$, we have that $\nabla v(\tau_1, \tau_2, B(\hat{x}))^\top \delta = 0$ for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all $\hat{x} \in \hat{\mathcal{X}}$ ($\mathcal{V}_S^{B(\hat{x})}$ from Definition 3.2.7 must be $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere zero-dimensional, if the TSM does not exhibit any USV state variables). We therefore have that, for $i = \bar{m} + 1, \bar{m} + 2, \dots, n$ and any $\tau_1, \tau_2 > 0$,

$$\frac{\partial \hat{v}(\tau_1, \tau_2, \hat{x})}{\partial \hat{x}^{(i)}} = \frac{\partial v(\tau_1, \tau_2, B(\hat{x}))}{\partial \hat{x}^{(i)}} = 0 \quad (3.45)$$

for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all $\hat{x} \in \hat{\mathcal{X}}$; in other words, the final \bar{p} entries of $\nabla v(\tau_1, \tau_2, B(\hat{x}))$ are equal to zero. This is intuitive: the final \bar{p} elements of the state process are unspanned, but because they are not USV state variables, bond covariances cannot have any sensitivity to them.

Using the process defined in Equation 3.42, and recalling that $\{\nabla \tilde{P}_t^T\}$ is $\lambda \otimes \mathbb{P}$ -almost surely

invertible, we have

$$(\nabla \tilde{P}_t^T)^{-1} \hat{v}^T(\hat{X}_t) (\nabla \tilde{P}_t^T)^{-\top} = \hat{\sigma}^Y(\hat{X}_t) \hat{\sigma}^Y(\hat{X}_t)^\top$$

$\lambda \otimes \mathbb{P}$ -almost surely. Both sides of this equations are functions of \hat{X}_t . Excluding when \hat{X}_t is in the $\hat{X}_*(\lambda \otimes \mathbb{P})$ -null subset of $\hat{\mathcal{X}}$ that we have excluded thus far, the derivative of the left-hand side with respect to the any of the final \bar{p} elements of \hat{X}_t is zero — this follows from Equations 3.45 and 3.44, and also the fact that $\nabla \tilde{P}_t^T$ depends only on Y_t . If we also exclude the $\hat{X}_*(\lambda \otimes \mathbb{P})$ -null values of \hat{X}_t where the above equation does not hold, we conclude that the right-hand side also has a partial derivative of zero. We thus have that, for $i = \bar{m} + 1, \bar{m} + 2, \dots, n$,

$$\frac{\partial}{\partial \hat{x}^{(i)}} \left(\hat{\sigma}^Y(\hat{x}) \hat{\sigma}^Y(\hat{x})^\top \right) = 0 \quad (3.46)$$

for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all $\hat{x} \in \hat{\mathcal{X}}$.

We then claim that, for $i = \bar{m} + 1, \bar{m} + 2, \dots, n$,

$$\frac{\partial}{\partial \hat{x}^{(i)}} \hat{\sigma}^Y(\hat{x}) = 0 \quad (3.47)$$

for $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost all $\hat{x} \in \hat{\mathcal{X}}$. In other words, we claim that the lack of sensitivity in the variance matrix given in Equation 3.46 implies the same lack of sensitivity in the diffusion matrix. We use the generalised lower-triangular form of $\hat{\sigma}^Y(\cdot)$ ensured by the first subsection of the proof to show this, and although we require a somewhat technical argument (that runs through the elements of a matrix of general size), the intuition is straightforward — consider, for instance

$$\begin{bmatrix} f(\hat{x}) & 0 \\ g(\hat{x}) & h(\hat{x}) \end{bmatrix} \begin{bmatrix} f(\hat{x}) & 0 \\ g(\hat{x}) & h(\hat{x}) \end{bmatrix}^\top = \begin{bmatrix} f(\hat{x}) & 0 \\ g(\hat{x}) & h(\hat{x}) \end{bmatrix} \begin{bmatrix} f(\hat{x}) & g(\hat{x}) \\ 0 & h(\hat{x}) \end{bmatrix} = \begin{bmatrix} f(\hat{x})^2 & f(\hat{x})g(\hat{x}) \\ f(\hat{x})g(\hat{x}) & g(\hat{x})^2 + h(\hat{x})^2 \end{bmatrix}.$$

The lack of sensitivity of the right-hand side can now be used, beginning with the first element of the first row:

$$0 = \frac{\partial}{\partial x^{(i)}} \left(f(\hat{x})^2 \right) = 2f(\hat{x}) \frac{\partial f(\hat{x})}{\partial x^{(i)}}, \quad (3.48)$$

which, provided that $f(\hat{x}) \neq 0$, implies that $\frac{\partial f(\hat{x})}{\partial x^{(i)}} = 0$. One can then verify that $\frac{\partial g(\hat{x})}{\partial x^{(i)}} = 0$, by considering an off-diagonal element (and recalling that $f(\hat{x}) \neq 0$ and $\frac{\partial f(\hat{x})}{\partial x^{(i)}} = 0$):

$$0 = \frac{\partial}{\partial x^{(i)}} \left(f(\hat{x})g(\hat{x}) \right) = f(\hat{x}) \frac{\partial g(\hat{x})}{\partial x^{(i)}} + g(\hat{x}) \frac{\partial f(\hat{x})}{\partial x^{(i)}}. \quad (3.49)$$

Finally, provided that $h(\hat{x}) \neq 0$, one can infer, from the lack of sensitivity in final diagonal element $g(\hat{x})^2 + h(\hat{x})^2$, that $\frac{\partial h(\hat{x})}{\partial x^{(i)}} = 0$. Importantly, the lower triangular construction does ensure that $f(\hat{x}) \neq 0$ and $h(\hat{x}) \neq 0$ ($\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere) — see Footnote 29.

To formally generalise this, firstly assume that the first m rows of the $\bar{m} \times m$ matrix $\hat{\sigma}^Y(\hat{x})$ are linearly independent, so that its rank-adding rows (the rows that contribute a dimension to the row-space, as it is increased from the first row downward) are its first m rows. Then the generalised lower triangularity described in the first subsection simply amounts to the first m rows of $\hat{\sigma}^Y(\hat{x})$

being lower triangular (see Equation 3.41). Note that

$$\begin{aligned}
[\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top]_{ij} &= \sum_{k=1}^m [\hat{\sigma}^Y(\hat{x})]_{ik} [\hat{\sigma}^Y(\hat{x})^\top]_{kj} \\
&= \sum_{k=1}^m [\hat{\sigma}^Y(\hat{x})]_{ik} [\hat{\sigma}^Y(\hat{x})]_{jk} \\
&= \sum_{k=1}^{\min(i,j,m)} [\hat{\sigma}^Y(\hat{x})]_{ik} [\hat{\sigma}^Y(\hat{x})]_{jk},
\end{aligned} \tag{3.50}$$

using lower triangularity ($[\hat{\sigma}^Y(\hat{x})]_{ik} = 0$ if $k > i$). From Equation 3.50, $[\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top]_{1,1}$ can be seen to equal $[\hat{\sigma}^Y(\hat{x})]_{1,1}^2$. We also have that $[\hat{\sigma}^Y(\hat{x})]_{1,1} \neq 0$ (for all but some $\hat{X}_*(\lambda \otimes \mathbb{P})$ -null \hat{x} ; see Footnote 29). For this first element, the reasoning used below Equation 3.48 shows that the property expressed in Equation 3.46 (namely, that $\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top$ is insensitive to its final \bar{p} arguments) extends to $\hat{\sigma}^Y(\hat{x})$ (as expressed in Equation 3.47).

Then consider $[\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top]_{2,1} = [\hat{\sigma}^Y(\hat{x})]_{2,1}[\hat{\sigma}^Y(\hat{x})]_{1,1}$, which — coupling with the first case, precisely as done below Equation 3.49 — shows that $[\hat{\sigma}^Y(\hat{x})]_{2,1}$ also shares the property that it is insensitive to its final \bar{p} arguments. Continuing along the first column of $\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top$ in this way shows that the relevant property is shared by the first column of $\hat{\sigma}^Y(\hat{x})$ (when considering the i th row, one gets $[\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top]_{i,1} = [\hat{\sigma}^Y(\hat{x})]_{i,1}[\hat{\sigma}^Y(\hat{x})]_{1,1}$, which shows that $[\hat{\sigma}^Y(\hat{x})]_{i,1}$ is insensitive to its final \bar{p} arguments).

Then consider working down the entries of the second column, beginning at the second entry ($[\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top]_{2,2}$), and then the third column, beginning at the third entry, and so on. When arriving at the i th element of the j th column in this fashion (so that $i \geq j$), one has that

$$[\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top]_{ij} = \sum_{k=1}^{\min(i,j,m)} [\hat{\sigma}^Y(\hat{x})]_{ik} [\hat{\sigma}^Y(\hat{x})]_{jk} = \sum_{k=1}^{\min(j,m)} [\hat{\sigma}^Y(\hat{x})]_{ik} [\hat{\sigma}^Y(\hat{x})]_{jk}. \tag{3.51}$$

Crucially, every entry involved on the far right-hand side except $[\hat{\sigma}^Y(\hat{x})]_{ij}$ has already been verified to share the key property (for $k < j$, the entries belong to columns checked before the j th is considered; for $k = j$, the extra term $[\hat{\sigma}^Y(\hat{x})]_{jj}$ has either been checked earlier in the column, or, if $i = j$, is the term of interest and will be squared). The lack of sensitivity on the far left-hand side can then be used to infer lack of sensitivity of $[\hat{\sigma}^Y(\hat{x})]_{ij}$, because we can assume that $[\hat{\sigma}^Y(\hat{x})]_{jj} \neq 0$ (this is because we have assumed that the first m rows of $\hat{\sigma}^Y(\hat{x})$ are independent — the reasoning in Footnote 29 then implies that the entries $[\hat{\sigma}^Y(\hat{x})]_{jj}$ are non-zero).

The assumption that the rank-adding rows of $\hat{\sigma}^Y(\hat{x})$ are its first m rows can be relaxed — instead of working through the first m columns of $\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top$, one needs to work through the columns corresponding to the rank-adding rows of $\hat{\sigma}^Y(\hat{x})$. A simple way to see this is to consider a re-indexing of the rows $\hat{\sigma}^Y(\hat{x})$ so that the first m rows (by the new index) are the rank-adding rows. With the variance matrix $\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top$ correspondingly re-indexed (from Equation 3.51, both the columns and rows need to follow the exact re-indexing applied to the rows $\hat{\sigma}^Y(\hat{x})$), the above argument can be applied — when lack of sensitivity to the final \bar{p} state variables is extended from the re-indexed $\hat{\sigma}^Y(\hat{x})\hat{\sigma}^Y(\hat{x})^\top$ to all of the entries of the re-indexed $\hat{\sigma}^Y(\hat{x})$, the re-indexing (which might need to depend on \hat{x}) can be undone and Equation 3.47 must hold.

With Equation 3.47 established, define a function $\tilde{\sigma}^Y : \mathcal{Y} \rightarrow \mathbb{R}^{\bar{m} \times m}$ with

$$\tilde{\sigma}^Y(y) = \hat{\sigma}^Y \left(\begin{bmatrix} y \\ \bar{u} \end{bmatrix} \right)$$

for all $y \in \mathcal{Y}$, where \bar{u} is a fixed element of \mathcal{W} (like that used in Lemma 3.1.17). Using precisely the same reasoning as in the proof of Lemma 3.1.17, it follows that³¹

$$\tilde{\sigma}^Y(Y_t) = \tilde{\sigma}^Y(\hat{X}_t).$$

□

Lemma 3.3.12. $\mu^Y(\cdot)$ can be written as a function of the first \bar{m} components of its argument; that is, there exists a function $\tilde{\mu}^Y(\cdot)$ such that

$$\tilde{\mu}^Y(Y_t) = \mu^Y(\hat{X}_t)$$

$\lambda \otimes \mathbb{P}$ -almost everywhere.

Remark 3.3.13. The intuition of the result is that the drift process of the spanned component is linked to the drift of ZCBs, which is given by $r_t P_{tT}$ (see Equation 2.6) and is therefore not affected by unspanned state variables $\{U_t\}$ (which do not affect any part of the term structure, including the short rate).

Note, however, that this only applies under a risk-neutral measure such as \mathbb{Q} . Under the real-world measure \mathbb{P} , ZCB drift (and therefore the drift process of the spanned component) *can* be influenced by $\{U_t\}$ (depending on the form of the market-price-of-risk process $\{\zeta\}$). If this is the case, some of the model's unspanned state variables play the role described in Remark 3.2.11.

Proof. Using Lemma 3.3.10, the dynamics of the spanned component in Equation 3.40 can be re-expressed as

$$dY_t = \mu^Y(\hat{X}_t) dt + \tilde{\sigma}^Y(Y_t) d\hat{W}_t^Y. \quad (3.52)$$

Define

$$m^Y(\tau, y) = -\frac{\partial \tilde{P}(\tau, y)}{\partial \tau} + \frac{1}{2} \sum_{i=1}^{\bar{m}} \sum_{j=1}^{\bar{m}} \frac{\partial^2 \tilde{P}(\tau, y)}{\partial y^{(i)} \partial y^{(j)}} [\tilde{\sigma}^Y(y) \tilde{\sigma}^Y(y)^\top]_{ij}$$

for all $\tau > 0$ and $y \in \mathcal{Y}$, so that the dynamics of a ZCB price process can be calculated with Itô's lemma, based on the reduced bond pricing function and Equation 3.52:

$$dP_{tT} = \left(m^Y(T-t, Y_t) + \nabla \tilde{P}(T-t, Y_t)^\top \mu^Y(\hat{X}_t) \right) dt + \nabla \tilde{P}(T-t, Y_t)^\top \tilde{\sigma}^Y(Y_t) d\hat{W}_t^Y.$$

In order to maintain Assumption 2.1.4 — in order for the cash-account-denominated ZCB price process to be a martingale — the above drift process must be equal to $r_t P_{tT}$. This, for any bond maturity T , gives an equation that holds for all $t \in [0, \min(T, S)]$. If we consider the \bar{m} maturity times engaged in the maturity process $\{\bar{T}_t\}$ (from the second subsection of the proof), we get the following vector equation:

$$m^{Y, \bar{T}}(Y_t) + \nabla \tilde{P}_t^{\bar{T}} \mu^Y(\hat{X}_t) = P_t^{\bar{T}} r_t, \quad (3.53)$$

where $m^{Y, \bar{T}}(Y_t)$ denotes a vector-valued process with typical element $[m^{Y, \bar{T}}(Y_t)]_i = m^Y([\bar{T}_t]_i, Y_t)$

³¹The difference between $\tilde{\sigma}^Y([y \ u]^\top)$ and $\tilde{\sigma}^Y([y \ \bar{u}]^\top)$ can be expressed by the line integral of $\nabla \tilde{\sigma}^Y(\cdot)$ along the straight curve between $[y \ u]^\top$ and $[y \ \bar{u}]^\top$. This line does not leave \mathcal{X} , and the line integral turns out to be equal to zero. This relies on the fact that the integrand is $\hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere zero (which follows from Equation 3.47 and the form of the eventual integrand — see Equation 3.20), and then on Assumption 3.0.2.

(also recall that $\{P_t^{\bar{T}}\}$ denotes the ZCB price processes, in a column vector, of the maturities given by the maturity process $\{\bar{T}_t\}$). Because $\{\nabla \tilde{P}_t^{\bar{T}}\}$ (based on the second subsection) is $\lambda \otimes \mathbb{P}$ -almost everywhere invertible, we then have

$$\mu^Y(\hat{X}_t) = (\nabla \tilde{P}_t^{\bar{T}})^{-1} (P_t^{\bar{T}} r_t - m^{Y, \bar{T}}(Y_t))$$

$\lambda \otimes \mathbb{P}$ -almost everywhere. Noting that the reduced bond pricing function can be used to express the bond prices $\{P_t^{\bar{T}}\}$ and the short rate (Equation 2.2 can be modified in this way), one sees that the right-hand side can be given by a function of the spanned component only. The function $\tilde{\mu}^Y(\cdot)$ can simply be defined based on this right-hand side, which proves the lemma. \square

Using Lemmas 3.3.10 and 3.3.12, we have that³²

$$dY_t = \tilde{\mu}^Y(Y_t) dt + \tilde{\sigma}^Y(Y_t) d\hat{W}_t^Y. \quad (3.54)$$

The spanned component $\{Y_t\}$ therefore, by virtue of satisfying this SDE, is adapted to the filtration generated by $\{\hat{W}_t^Y\}$, which we denote $\{\hat{\mathcal{F}}_t^Y\}$.

Remark 3.3.14. Remark 3.2.12 discusses the case that USV state variables are present, but are unaccompanied by USV factors. One would not be able to reduce the diffusion function as far as is done in Lemma 3.3.10 — Equation 3.46 would not necessarily hold for all of the final \bar{p} state variables.³³

Suppose an \mathcal{S} -transform can be performed such that the first $n - \bar{p} + p^{\text{SV} - \sigma}$ rows of $J_A(x)$ span $\text{Im}(\sigma(x))^\perp$, where $p^{\text{SV} - \sigma}$ is the dimension of $\text{Im}(\sigma(x))^\perp \cap \mathcal{S}^x$ (see Equation 3.33 — here we must also assume that this dimension does not vary across the state space). This is a slightly stronger construction than that given in Remark 3.2.12 — it ensures that the relevant rows of $J_A(x)$ lie in the cokernel of $\sigma(x)$, so that $p^{\text{SV} - \sigma}$ of the unspanned state variables are locally deterministic, and also that Equation 3.46 holds for the other unspanned state variables (see Footnote 33 and note that, given this construction, $\bar{p} - p^{\text{SV} - \sigma}$ of the final \bar{p} columns of $J_B(A(x))$ must lie in the image of $\sigma(x)$).

Suppose also that the case of *weak USV* outlined in Remark 3.2.12 is avoided — that is, suppose that the drift functions of the locally deterministic unspanned state variables do not depend on the other unspanned state variables. Then consider appending the locally deterministic unspanned state variables onto the spanned component. The resulting dynamics would form a valid SDE based on $\{\hat{W}_t^Y\}$ (implying that this larger set of state variables will be adapted to $\{\hat{\mathcal{F}}_t^Y\}$). This is because, firstly, the spanned state variables' diffusion function can be reduced to depend only on themselves and the locally deterministic unspanned state variables, and the additional diffusion function is equal to zero; secondly, Lemma 3.3.12 still applies to the spanned component's drift function (and the drift of the locally deterministic unspanned state variables is assumed to not depend on the other unspanned state variables). One can then proceed to the final subsection of the proof.

³²The $\lambda \otimes \mathbb{P}$ -null difference between $\{\mu^Y(\hat{X}_t)\}$ and $\{\tilde{\mu}^Y(Y_t)\}$ will only cause \mathbb{P} -null (and therefore \mathbb{Q} -null) differences to Y_t at any $t \in [0, S]$, and can thus be overlooked — see Footnote 10. This idea is used in the final subsection of the proof, where the $\lambda \otimes \mathbb{P}$ -almost everywhere invertibility of $\{\nabla \tilde{P}_t^{\bar{T}} \sigma^Y(\hat{X}_t)\}$ is sufficient, as the $\lambda \otimes \mathbb{P}$ -null part of the integrand where it is not invertible can be ignored.

³³This is because, without the lack of USV state variables, one would not have $\nabla v(\tau_1, \tau_2, B(\hat{x}))^\top \delta = 0$ for all $\delta \in \mathcal{S}^{B(\hat{x})}$, but only for $\delta \in (\mathcal{S}^{B(\hat{x})} \cap \text{Im}(\sigma(x)))$, which is implied by the lack of USV factors.

Attainability of a General Claim

Consider now a T^* -claim Θ . We assume the claim is $\mathcal{F}_{T^*}^P$ -measurable and bounded in the sense of Equation 2.11, ensuring that

$$\hat{M}_t^Y = \mathbb{E}^\mathbb{Q}[D_{T^*}\Theta \mid \hat{\mathcal{F}}_t^Y]$$

is well-defined for all $t \in [0, T^*]$. It is crucial to note that $\mathcal{F}_{T^*}^P \subseteq \hat{\mathcal{F}}_{T^*}^Y$, because ZCBs can be given by the spanned component only (via the reduced bond pricing function), which is adapted to $\{\hat{\mathcal{F}}_t^Y\}$. This implies that $\hat{M}_{T^*}^Y = D_{T^*}\Theta$ (because $D_{T^*}\Theta$ is $\mathcal{F}_{T^*}^P$ -measurable, and is therefore also $\hat{\mathcal{F}}_{T^*}^Y$ -measurable).

The process $\{\hat{M}_t^Y\}$ is clearly a $(\mathbb{Q}, \{\hat{\mathcal{F}}_t^Y\})$ -martingale. The martingale representation theorem asserts that there exists an adapted, d -dimensional process $\{\hat{\Gamma}_t\}$ such that³⁴

$$\hat{M}_t^Y = \hat{M}_0^Y + \int_0^t \hat{\Gamma}_u^\top d\hat{W}_u^Y.$$

Suppose we construct a portfolio based on the maturity process $\{\mathcal{T}_t\}$ from the second subsection of the proof (which ensures the invertibility used below) and on a holdings process $\{\phi_t\}$ defined, for all $t \in [0, T^*]$, with

$$\phi_t = (\nabla \tilde{P}_t^\top \sigma^Y(\hat{X}_t))^{-\top} \frac{\bar{\epsilon}(\hat{X}_t) \hat{\Gamma}_t}{D_t}.$$

Modifying Equation 2.8 to reflect bond dynamics based the reduced bond pricing function and Equation 3.38, the resultant portfolio's discounted value process has dynamics

$$\begin{aligned} d(D_t V_t^{\mathcal{T}, \phi}) &= D_t \phi_t^\top \nabla \tilde{P}_t^\top \sigma^Y(\hat{X}_t) dW_t^Y \\ &= \frac{D_t \hat{\Gamma}_t^\top}{D_t} \bar{\epsilon}(\hat{X}_t)^\top (\nabla \tilde{P}_t^\top \sigma^Y(\hat{X}_t))^{-1} \nabla \tilde{P}_t^\top \sigma^Y(\hat{X}_t) dW_t^Y \\ &= \hat{\Gamma}_t^\top \bar{\epsilon}(\hat{X}_t)^\top dW_t^Y \\ &= \hat{\Gamma}_t^\top d\hat{W}_t^Y. \end{aligned}$$

Setting the initial capital of the portfolio $V_0^{\mathcal{T}, \phi}$ to \hat{M}_0^Y , it follows that $D_{T^*} V_{T^*}^{\mathcal{T}, \phi} = \hat{M}_{T^*}^Y$, and then that $V_{T^*}^{\mathcal{T}, \phi} = \Theta$. This shows that the TSM in question satisfies economic completeness as per Definition 2.2.6, completing the proof of Theorem 3.2.10.

3.4 Ensuing Remarks and Examples

3.4.1 Partially Spanned Factors

Recall that there are two ways for a vector to be included in the unspanned factors direction set \mathcal{U}^x in Definition 3.1.2 (see the discussion in the proof of Lemma 3.1.18): it can correspond to a direction of state process movement that has no effect on the term structure, or it can lie in the kernel of the diffusion process.

It is not clear whether a partially spanned factor (Definition 3.1.8) can arise from directions of the former type; the author is not aware of any TSM with an unspanned factor direction set

³⁴See Filipović [2009, Ch.4.4] or Karatzas & Shreve [2012, Ch.3.4]. Note that the extension described in Footnote 20 is not necessary here, because the Brownian motion used as the integrator is the same one that generates the martingale's filtration.

that has different dimension over $X_*(\lambda \otimes \mathbb{P})$ -non-null parts of the state space (modulo the latter, kernel-based unspanned factors directions). Sections 3.1 and 3.2 therefore raises the question of whether such models exist, and if so, whether they have any practical or empirical merit.

Constructing a model with a partially spanned factor, based on the kernel-type of unspanned factors directions, is relatively easy. In a two-dimensional setting, a state process constituted by the short rate $\{r_t\}$ and a mean-reversion process $\{\theta_t\}$ could satisfy

$$d \begin{bmatrix} r_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} \kappa_r(\theta_t - r_t) \\ \kappa_\theta(\theta_t - \theta_t) \end{bmatrix} dt + \begin{bmatrix} \sigma_r(r_t) & 0 \\ 0 & \sigma_\theta(r_t, \theta_t) \end{bmatrix} dW_t$$

for constants κ_r , κ_θ and θ_θ , and real-valued functions $\sigma_r(\cdot)$ and $\sigma_\theta(\cdot, \cdot)$, with $\sigma_\theta(\cdot, \cdot)$ equal to zero on a $X_*(\lambda \otimes \mathbb{P})$ -non-null set $\mathcal{X}^* \subseteq \mathcal{X}$. While the TSM would contain no unspanned state variables, for $x \in \mathcal{X}^*$, one finds that³⁵

$$\mathcal{U}^x = \left\{ \begin{bmatrix} 0 & \alpha \end{bmatrix}^\top \mid \alpha \in \mathbb{R} \right\},$$

and $\dim(\mathcal{U}^x) = 0$ elsewhere. Assuming that $\mathcal{X} \setminus \mathcal{X}^*$ is also $X_*(\lambda \otimes \mathbb{P})$ -non-null, the TSM would exhibit one partially spanned factor according to Definition 3.1.8. Although $\sigma_\theta(\cdot, \cdot)$ would need to obey some constraints (such as differentiability in Assumption 2.1.5, or, more generally, the Lipschitz and linear growth conditions), it could easily be specified in a way to achieve this, and also to admit some qualitative interpretation. For example, $\sigma_\theta(\cdot, \cdot)$ could be constructed so that it is equal to zero if and only if the sum of the short rate and mean-reversion level are less than a certain amount — the resultant model would thus be a one-factor model in this low-interest rate regime, but a two-factor model elsewhere.

Such models, suggested by Sections 3.1 and 3.2, are not present in the literature (to the author's knowledge). Future research could interrogate how such models fare empirically, particularly if one can link observed economic circumstances to the qualitative nature of the specification (for instance, the above example specification of $\sigma_\theta(\cdot, \cdot)$ could be implemented over a period where a low interest-rate regime is entered).

One could also consider extending our completeness results to these models. They would certainly be mathematically incomplete, because a non-trivial kernel of the diffusion process $\sigma(x)$ (over a $X_*(\lambda \otimes \mathbb{P})$ -non-null part of the state space) means that some part of the Brownian motion's variation is not being transmitted to the state process, and therefore cannot be hedged with ZCBs (in the above example, whenever $X_t \in \mathcal{X}^*$, the second Brownian motion is not affecting the model at all, so any claims depending on $W_t^{(2)}$ at these times will not be attainable). However, because this part of the Brownian motion is not affecting ZCBs, the model may nevertheless be economically complete — perhaps the relevant parts of the Brownian motion can be separated from the ZCBs, like the fully unspanned factors are separated in the proof of Theorem 3.2.10.

3.4.2 Time-inhomogeneous Models

Here we consider relaxing the time homogeneity imposed in Assumption 3.0.1. Without time homogeneity, the theory of Sections 2.1 and 2.2 must, *in principle*, be applied to a TSM at all

³⁵No unspanned state variables would be exhibited, firstly because, this tends not occur without specific and non-trivial constraints on the TSM, and secondly because, in this case, $\{r_t\}$ and $\{\theta_t\}$ would give the level of the short and long end of the yield curve respectively, and so bond prices of any (non-zero and finite) maturity would need to depend on both of these values.

calendar times t ; for instance, unspanned factor directions *at time t* are given by

$$\mathcal{U}^{x,t} = \{\delta \in \mathbb{R}^d \mid \nabla P(t, t + \tau, x)^\top \sigma(t, x) \delta = 0 \quad \forall \tau > 0\}.$$

However, this separate consideration of each calendar time is typically unnecessary — the same results will usually be obtained for all $t \in [0, S]$, because the time dependency of a TSM is usually a calibration-motivated *shift* of the state process or bond pricing function, rather than a time dependence in the model’s fundamental characteristics.³⁶

Time dependency is in fact usually used for the specific purpose of calibrating the TSM to the observed initial term structure — such models can often be viewed as a time-homogeneous model with a deterministic, time-dependent shift:

$$P_{tT} = P^C(t, T) P^{\text{TH}}(T - t, X_t), \quad (3.55)$$

where $P^{\text{TH}}(\cdot, \cdot)$ is $\mathcal{C}^{1,2}$ a function that, in addition to a time to maturity, takes the value of some time-homogeneous state process (satisfying Equation 3.2), and $P^C(\cdot, \cdot)$ is a deterministic function that allows for calibration. An example is the famous Hull & White [1990] model, which can be viewed, Brigo & Mercurio [2007] show, as per Equation 3.55.^{37,38} Other examples can be found amongst Heath et al. [1992] (HJM) models — if an HJM model has a *time-homogeneous volatility specification* (that is, the forward rate’s diffusion function does not depend on calendar and maturity time separately) and can be given a *finite-dimensional state process representation* (discussed below), it can be written in the form of Equation 3.55.

If a time-inhomogeneous TSM admits the representation indicated by Equation 3.55, the theory of Sections 3.1 and 3.2 is valid to apply to the time-homogeneous part of the bond pricing function $P^{\text{TH}}(\cdot, \cdot)$ — the results, including the proposed transformations and the completeness theorems, hold, with only minor notational adjustments (such as the inclusion of the time-dependent function $P^C(\cdot, \cdot)$ where appropriate). The HJM model constructed in Trolle & Schwartz [2009] is an example of such a model. For some natural number N , it involves $2N$ factors, N of which are USV factors. An $8N$ -dimensional state process is required for the bond prices to be represented as per Equation 3.55.

The possibility of a finite-dimensional state process representation for HJM models (which are specified with an infinite number of state variables) has been thoroughly investigated (Björk & Svensson [2001] and Chiarella & Kwon [2003] are the culmination of this literature). These models are in fact well-suited to our distinction between factors and state variables because — as seen in the Trolle & Schwartz [2009] model mentioned above — many more state variables than factors are typically required for this representation.

A possible direction for future research is a generalisation of Section 3.1 and 3.2 to a framework that does not assume a finite-dimensional state process. While this would pose many challenges, one significant benefit would be the possibility of circumventing the need for an invertible state process transformation.

³⁶One *could* construct a model that changes over time in a fundamental way, but the motivation for such a model would not be clear. For example, consider a short-rate process satisfying

$$dr_t = \kappa(\theta - r_t) dt + \bar{\sigma}(t, r_t) dW_t,$$

where $\bar{\sigma}(\cdot, \cdot)$ is a function with the property that $\bar{\sigma}(t, r) = 0$ for all $t > \bar{T}$. The resultant model would then have no unspanned factors for $t \in [0, \bar{T}]$, and would gain an unspanned factor after \bar{T} .

³⁷This is generalised by Brigo & Mercurio [2001] — a method of adding a time dependency to time-homogeneous models in the vein of Equation 3.55 (so that they can be calibrated to an initial term structure) is proposed.

³⁸In Section 4.1.2, we use the Hull & White [1990] model in this form, i.e., a form where the time dependency is explicitly and separately reflected.

Finally, note the following simple way in which time-dependency can be captured without relaxing Assumption 3.0.1: one could allow the first state variable to evolve deterministically along with time (i.e., set $X_t^{(1)} = t$). Checking the various inhomogeneous time points would then amount to working through the enlarged state space (as Definitions 3.1.1 and 3.1.2 require). A disadvantage of the approach would be that a further state variable, relative to the number of factors, would be required, complicating the results.

3.4.3 Non-Brownian Cases

Relaxing the assumption of a Brownian-based state process would give our theory a very wide applicability. Although the distinction between state variables and factors may be less clear or invalid in non-Brownian frameworks, our definition of unspanned state directions (Definition 3.1.1) is certainly suitable in this more general setting. Our proposed method of counting unspanned state variables (based on the maximum and minimum dimension of unspanned state directions, up to $X_*(\lambda \otimes \mathbb{P})$ -nullity) links to the completeness of Brownian TSMs in a way that suggests this is a suitable method whether or not completeness is at stake. The number of unspanned state variables (from Definitions 3.1.12 and 3.1.13) are validly interpreted as the number of dimensions, out of the total n dimensions, of state process movements that have no effect on the term structure.

Whether such an unspanned state variable plays the role of introducing stochastic volatility can similarly be determined with Definition 3.2.9 (and the earlier definitions on which it depends), provided a suitable generalisation to the bond covariance function can be found. Equation 3.30 suggests this generalisation, assuming the state process $\{X_t\}$ is Markovian:

$$v(\tau_1, \tau_2, X_t) = \lim_{\delta \downarrow 0} \frac{1}{\delta} \mathbb{C}\mathbb{V}^{\mathbb{P}}[P(\tau_1 - \delta, X_{t+\delta}), P(\tau_2 - \delta, X_{t+\delta}) | \mathcal{F}_t]. \quad (3.56)$$

If this limit exists and results in a differentiable function of the state process, our definitions can be applied. Suppose, for instance, that the state process is a jump diffusion satisfying

$$dX_t = \mu(X_t) dt + \sigma(X_t) dW_t + \sigma_J(J^{(t)}) dN_t,$$

where Equation 3.2 has been embellished with an \mathbb{R}^n -valued function $\sigma_J(\cdot)$, a set of (independent and identically distributed) jump random variables $\{J^{(t)}\}$ and a Poisson process $\{N_t\}$. From Crépey [2013, Ch.3.2], the limit in Equation 3.56 does exist in this case and is given by

$$\begin{aligned} v(\tau_1, \tau_2, x) &= \nabla P(\tau_1, x)^\top \sigma(x) \sigma(x)^\top \nabla P(\tau_2, x) \\ &\quad + \mathbb{E}^{\mathbb{P}} \left[\left(P(\tau_1, x + \sigma_J(J^{(t)})) - P(\tau_1, x) \right) \left(P(\tau_2, x + \sigma_J(J^{(t)})) - P(\tau_2, x) \right) \right]. \end{aligned}$$

The second term indicates an additional way in which bond covariances are affected by the state variables. Whether USV models can be constructed in this vein is a potential avenue for future research.

3.4.4 Unspanned Stochastic Correlation

Movements of the state process in any USV state direction (Definition 3.2.7) affect some aspect of the bond covariance structure, but Section 3.2 does not depend on what aspect this is. In practice, one finds that the whole covariance structure is sensitive to these movements — in Example 4, for instance, Equation 3.32 shows that changing the second state variable (which moves the state process in USV state directions) affects bond covariances for any pair of maturities τ_1 and τ_2

(rather than only for certain pairs). This is the case for the USV models introduced in the final subsection of Section 4.1.2 below, and in fact for all existing USV models that conform to the framework of Section 2.1.³⁹

This raises the question of whether counterexamples exist. In particular, the discussion below Definition 3.2.2 alludes to a specific type of counterexample that separates volatility and correlation. If, for example, moving the state process in an unspanned direction affects bond covariance at certain maturity pairs (i.e., changes the value of $v(\tau_1^*, \tau_2^*, x)$ for some maturities τ_1^* and τ_2^*) but does not affect bond volatility (that is, does not affect $v(\tau, \tau, x)$ for any τ), the direction could be termed an *unspanned stochastic correlation (USC)* direction. If bond volatility is affected, the direction is not a USC direction, so can be termed a *strict USV* direction.

To capture this formally, first note that

$$\{\delta \in \mathcal{V}_{\mathcal{U}}^x \mid \nabla v(\tau, \tau, x)^\top \sigma(x) \delta = 0 \quad \forall \tau > 0\}$$

gives a subspace of the unspanned factor directions $\mathcal{V}_{\mathcal{U}}^x$ whose elements are USC factor directions: they do not affect volatilities, but do affect covariances of distinct maturities (by virtue of being in $\mathcal{V}_{\mathcal{U}}^x$) and therefore correlations. Then we can define the *strict USV factor directions*, for each $x \in \mathcal{X}$, with

$$\mathcal{V}_{\mathcal{U}, \text{vol}}^x = \{\delta \in \mathcal{V}_{\mathcal{U}}^x \mid \nabla v(\tau, \tau, x)^\top \sigma(x) \delta = 0 \quad \forall \tau > 0\}^\perp,$$

where $\{\cdot\}^\perp$ denotes the orthogonal complement with respect to $\mathcal{V}_{\mathcal{U}}^x$. The difference between the dimension of $\mathcal{V}_{\mathcal{U}}^x$ and of $\mathcal{V}_{\mathcal{U}, \text{vol}}^x$ therefore gives the number of USC factors prevailing at that point. The procedure used throughout Sections 3.1 and 3.2 must then be used: the state space is partitioned according to the dimension of $\mathcal{V}_{\mathcal{U}, \text{vol}}^x$, and the highest-dimensional block that is $X_*(\lambda \otimes \mathbb{P})$ -non-null is the number of strict USV factors. The difference between this and number of USV factors from Definition 3.2.5 then gives the number of USC factors.

In addition to constructing USC models, future research could interrogate them empirically (for example by addressing the relative pricing of caps and swaptions mentioned in Footnote 39).

³⁹Collin-Dufresne & Goldstein [2001] propose a TSM where correlations have their own source of stochastic variation that cannot be hedged with ZCBs, but the model is infinite-dimensional and thus does not conform to our framework. They suggest that this kind of stochastic correlation — which we term unspanned stochastic correlation — can resolve the apparent mismatch between the cap and swaption markets that they and others (such as Longstaff, Santa-Clara & Schwartz [2001]) document.

Chapter 4

Empirical USV Evidence

The chapter begins by reviewing the empirical USV literature. Section 4.1.1 starts with what we call the *original evidence* of Collin-Dufresne & Goldstein [2002] and Heidari & Wu [2003] (hereafter CG and HW, respectively), and then considers the use of hedging to ascertain whether USV models are necessary. This approach’s merits are discussed, as are the specific studies of Fan et al. [2003] and Li & Zhao [2006]. Our specific method, including our data and benchmark models, is detailed in Section 4.1.2. Section 4.2 presents the results in a number of subsections, and Section 4.3 concludes.

4.1 Empirical Approach

4.1.1 Evidence Review

Original Evidence

CG calculate, from their raw data, monthly returns based on holding a cap-floor straddle: a combination of a long cap and long floor, both, in this case, at-the-money. The positive and negative exposure to interest rates, in the cap and floor respectively, roughly cancel out, but a positive exposure to volatility remains — straddle returns are large when interest-rate movements, whether positive or negative, are large. The straddle returns are regressed against contemporaneous changes in swap rates of several maturities. Low R^2 s — mostly between 0.15 and 0.4 — are found. Citing this apparent mismatch between the underlying and derivative markets, CG conclude that there are additional sources of variation driving cap and floor prices, conceiving and indeed coining the notion of USV.

Secondarily, CG also adduce the fact that their regression residuals are highly cross-correlated and have a dominant principal component (explaining around 85% of their total variance). Rather than being due to misspecification noise or market microstructure, they suggest that their residuals are systematically driven, and can be ascribed to a small number of factors responsible for the observed underlying-derivative market mismatch.

HW’s approach is conceptually very similar: a derivative-market quantity (in this case, at-the-money swaption-implied volatilities) is regressed against underlying-market quantities (the three leading yield-curve principal components). An important methodological difference, though, is HW’s focus on *levels* as opposed to *changes* (or returns): implied volatility levels are regressed against principal components based on yield levels (since the famous work of Litterman & Scheinkman [1991], principal components have been calculated on both yield levels and changes

— see Piazzesi [2010, §7.2]). HW interpret R^2 s of less than 0.6 as evidence of derivative-market variation that is unspanned by the underlying ZCB market.

Bikbov & Chernov [2009] criticise the original evidence of CG and HW for having no prior sense of what R^2 values would warrant a USV conclusion. They, amongst others beginning with Fan et al. [2003], note that R^2 s being less than one could simply be the result of a misspecified regression, rather than an indication of a structural difference between the underlying and derivative markets. Indeed, these regression tests are heuristic in nature — no rigorous justification is given for the posited linear relationship. To interrogate this, Bikbov & Chernov [2009] simulate data from both a USV and standard TSM, and regress the simulated implied-volatility series against simulated yields. They thus consider the levels-based approach of HW, which, they argue, makes their study more prudent from a USV-sceptic point of view — regressions based on levels tend to have higher R^2 s than those based on changes, and this makes it harder to reach a USV-positive conclusion.¹ Bikbov & Chernov [2009] find that the estimated R^2 s do not differ significantly when the underlying model is changed, indicating that a level-based regression of this sort is not an effective device for discriminating USV from non-USV data.

Any method that purports to address the USV empirical question can be assessed in this way: data can be simulated from USV and from standard models, and one can check whether the method *correctly identifies the presence and absence of USV*.² We will apply this critique — which we refer to as *benchmarking* — to CG’s changes-based regressions, as well as to *hedging-based methods*, which assess whether, according to historical data, volatility-sensitive derivatives can be hedged effectively with underlying instruments such as ZCBs.³ We view this as a natural generalisation of the original evidence’s regression approach.

A Hedging Approach

The changes-based regressions of CG are direct tests of the possible underlying-derivative market disconnect implied by USV models: derivative-market variation is related to underlying-market variation, and the degree of observed relatedness is quantified. Implementing hedges in the interest-rate market generalises this: derivative price changes are linked to ZCB price changes, and the link is allowed to be more flexible — instead of a constant regression coefficient, ZCB price changes are scaled by model-implied hedge ratios, which depend on the state of the world (such as the level of interest rates at that time).

Such hedges assess the link between the derivative and underlying markets in the fundamentally relevant way: whether derivative-price changes correspond to ZCB-price changes in some way (or whether a portion of the derivative-price changes is independent) is the central issue at stake in the empirical USV literature. This is the first reason behind this chapter’s focus on a hedging approach. The second is that the two hedging-based studies in the empirical USV literature — due to Fan et al. [2003] and Li & Zhao [2006] — arrive at starkly conflicting conclusions. We will attempt to resolve the apparent conflict with the benchmarking idea described above. Thirdly,

¹The persistence of levels-based time series can inflate estimated correlations. The spurious regression problem — see Engle & Granger [1987], or Murray [1994] for an intuitive explanation — for non-stationary time series is well known, and a stationary, persistent series can, over a finite sample, give spurious correlation estimates.

²Phrases such as “the presence of USV factors” or “data with USV” are common in the USV literature. Strictly speaking, this usage is inaccurate — USV is a feature of certain models, not of datasets or markets. It is nevertheless a useful shorthand, and we will continue to “test for the presence of USV”, where we are in fact assessing whether the data in question appears to have come from a USV model.

³Although not emphasised in their paper, and not addressed by Bikbov & Chernov [2009], CG do perform a simulation check of this sort. Their goal in doing so is to verify that their monthly frequency is inconsequential, rather than to interrogate the regression’s linear relationship, which may explain why Bikbov & Chernov [2009] do not address their finding. CG conclude that their changes-based regression can discriminate well, with an R^2 s decreasing from 0.95, in the non-USV case, to 0.25 when the simulations come from a USV model. We attempt to replicate and extend this finding in Section 4.2.1.

given the immense importance of hedging interest-rate derivatives, it is interesting and relevant to interrogate how the possible presence of USV manifests at the level of hedging effectiveness.⁴

Furthermore, we find that the weaknesses of a hedging approach — particularly the model dependence of the hedge ratios — are well managed. Although a model needs to be selected somewhat arbitrarily, and any choice involves misspecification, the hedge ratios from even a simple model bring considerable improvement over constant regression coefficients. In Section 4.2, our benchmark simulations results will show that this improvement is sufficient for the link between the underlying and derivative markets to be suitably assessed.

Hedging-based USV Evidence

Fan et al. [2003] (FGR hereafter) implement one-, two-, three- and four-factor Heath et al. [1992] (HJM) models in the LIBOR market. The covariance structure is determined partly by historically estimated covariances and partly by cross-sectional calibration to at-the-money swaptions. The HJM model informs ZCB hedge strategies for these swaptions, rebalanced at a weekly frequency. The results are reported in terms of *hedging* R^2 s, defined with

$$R^2 = 1 - \frac{SS_{\text{hedged}}}{SS_{\text{unhedged}}},$$

where SS_{hedged} denotes the sum of the squared hedge residuals, and SS_{unhedged} the sum of the squared changes in the (unhedged) swaption price. FGR found their hedging to be effective, with hedging R^2 s in the region of 0.9 with the three- and four-factor model.

FGR conclude that any USV-related variance they are unable to hedge is very minor, considering that their hedge residuals certainly include a degree of time-discretisation and model-misspecification error. Secondly, they note that there is no dominant principal component amongst hedging residuals (the leading entry accounts for only 42% of the total residual variance) — rather than being systematically driven by USV, the hedge residuals resemble noise. As a further secondary test, FGR regress their hedging residuals on changes in (swaption-)implied volatilities and find (regression) R^2 s no larger than 0.6. As many volatilities are included, they also report adjusted R^2 s, which are less than 0.4, and take this as confirmation of their negative-USV conclusion — changes in implied volatilities (presumably a proxy for changes in a possible USV state variable) do not seem to strongly explain the (minor) portion of swaption price variation they were unable to hedge.

Li & Zhao [2006] (LZ hereafter) use a three-factor quadratic term structure model to hedge LIBOR-written caps. Their hedging methodology is peculiar. The core literature that empirically hedges interest-rate derivatives — notably Driessen, Klaassen & Melenberg [2003], Gupta & Subrahmanyam [2005] and Fan, Gupta & Ritchken [2007] — as well as FGR, abide by the following methodological tenets:

1. The hedging models are calibrated, at each rebalancing, to the market-prevailing yield curve. In other words, time-inhomogeneous models are used.
2. At each rebalancing, at least some parameters are determined so as to match (perhaps approximately) prevailing cap or swaption prices.

⁴We note that we do not address Li & Zhao [2009], who non-parametrically estimate distributions implicit in prevailing cap prices. This approach is very different to ours, and we view it as complementary. We also note that Collin-Dufresne et al. [2009] and Andersen & Benzoni [2010] are concerned with affine term structure models only and are not addressed in this thesis.

3. Yield-curve-based instruments, such as ZCBs or swaps, are engaged according to the model-identified hedge portfolio, which, because of the calibration in tenet 1, can in fact be purchased at the model-reflected price.

LZ follow a different course with regard to each point above, namely:

1. The quadratic term structure model is used in its time-homogeneous form and the model is therefore not correctly calibrated to prevailing yield curves.⁵
2. Parameters are determined via quasi-maximum likelihood based on an extended Kalman filter. This filter *runs over the whole hedging period* and therefore makes use of future information that one does not have in a true hedging exercise.⁶
3. The filtered state variables — the Kalman filter’s estimate for the location of the three state variables at each cross section — *are assumed to be tradeable and are used as hedging instruments*.

LZ defend the third point by pointing out that this favourable hedging condition makes their study more prudent, in that it should make it harder to identify USV in their hedging residuals. Perhaps the second point could be defended in a similar way.

In spite of this, LZ find their hedging of caps to be ineffective, with hedging R^2 s in the region of 0.6 (although there is significant variation across the various strikes and maturities they consider). Contrasting this against very effective *bond hedging* — i.e., the hedging of a ZCB with, in their case, the filtered state variables — (with R^2 s around 0.95) and the good fit the quadratic term structure model appears to have with the time series of bond yields, LZ conclude that USV is present in their data.

In a fashion similar to FGR, LZ also conduct some secondary tests. They identify the leading principal components of their yield curve data and use the changes to attempt to explain their hedging residuals in a linear regression. This regression has a very poor explanatory power (with an R^2 of less than 0.1), but improves significantly (with R^2 s between 0.21 and 0.95 depending on the particular cap underlying the hedge) when regressors include changes in the first four principal components of at-the-money cap-implied volatilities. They adduce this as evidence of USV — the residuals seem to be linked to volatility changes rather than imperfectly hedged movements in the primary market. Finally, they do not find a dominant principal component amongst their hedge residuals and conclude that multiple USV factors are likely present.

It is worth noting that both FGR and LZ hedge straddles (a swaption straddle, in FGR’s case, is a combination of a long payer and long receiver swaption) in addition to caps/swaptions.⁷ This deteriorates their hedging effectiveness — FGR’s average R^2 changes from about 0.9 to about 0.8, while some of LZ’s R^2 s drop to less than 0.2. This is despite gamma hedges that both studies put in place, because of the low first-order sensitivities of straddles.⁸ For reasons explained in the next subsection, we will refrain from hedging straddles and therefore from gamma hedging.

⁵In their introduction, LZ criticise FGR for choosing models that take an initial yield curve as exogenous, claiming that FGR therefore do not answer the key question of whether models “can price both bonds and interest-rate derivatives”. But these models are simply used to provide hedge ratios that make sense in light of the relevant hedge instrument costs. What matters is not the source of the hedge ratios, but whether hedge ratios exist that can consistently offset the variation in the derivatives with variation in underlying instruments.

⁶The historical covariances used by FGR are estimated over a period prior to their hedging period, and thus do not make use of future information

⁷Some of the hedging residuals used in the secondary regressions mentioned above are from straddle hedges.

⁸It is not clear, however, how LZ conduct gamma hedging — they assume the state variables can be traded directly, which does not offer a way of attaining second-order sensitivity (like, for example, how an underlying stock is unable to hedge an equity option’s gamma).

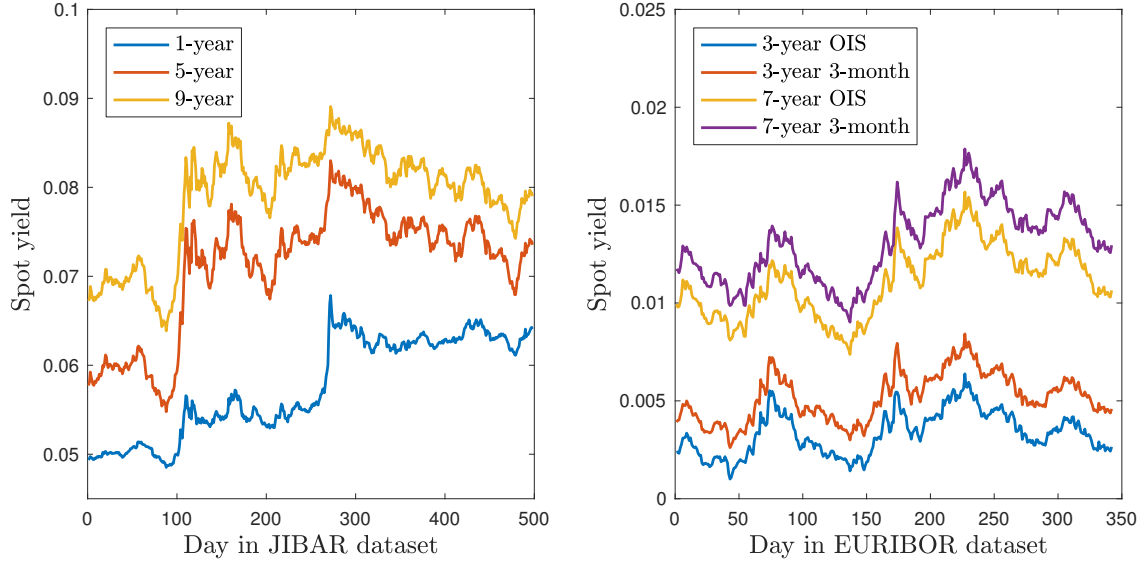


Figure 4.1: Time series of selected (continuously compounded) yields, over the sample periods of either dataset.

We should also note the necessary *asymmetry* in the USV-related interpretation of hedging results — if hedging is found to be effective, there cannot be significant presence of USV; ineffectual hedging, on the other hand, *might* indicate unspanned stochastic volatility risk in the derivative market, or it might be due to other problems in the hedging implementation, such as a misspecified model.

In summary, the two hedging-based studies in the USV empirical literature have conflicting results and conclusions. Because of the interpretation asymmetry, FGR’s negative conclusion is arguably the stronger of the two, despite it conflicting with the majority of the literature. This, as well as LZ’s unorthodox methodology, suggest that FGR’s work should not be overlooked.

We refer to the tests that follow the basic hedging results — mostly regressions of the hedge residuals — as *secondary tests*. These, in combination with benchmark simulations, will be instrumental in resolving the FGR-LZ conflict in Section 4.2.

4.1.2 Data and Methodology

Data

In an effort to make our conclusions as general as possible, we work with two datasets, each from markets which are different in nature: one is emerging with relatively high interest rates, while one is highly developed with low rates in the relevant period.

The first is based on caps written on JIBAR (the Johannesburg Interbank Agreed Rate). For each of the 498 trading days in 2013 and 2014, we have a bootstrapped yield curve, and, for two strike rates of 6% and 7%, a term structure of ten implied volatilities. The volatilities are converted, using Black’s formula and the relevant yield curve, into cap prices. The first of the ten caps, of either strike, is comprised of four quarterly caplets with pay out dates in three, six, nine and twelve months’ time. The second cap includes eight such adjacent, quarterly caplets, and so on, up to the tenth cap which is comprised of forty caplets. Taking the difference between each consecutive cap price we get, for both strikes, ten *difference cap* prices. Each difference cap is comprised of four caplets; the 6-year difference cap, for instance, involves four quarterly caplets with pay outs beginning in five and a quarter years.

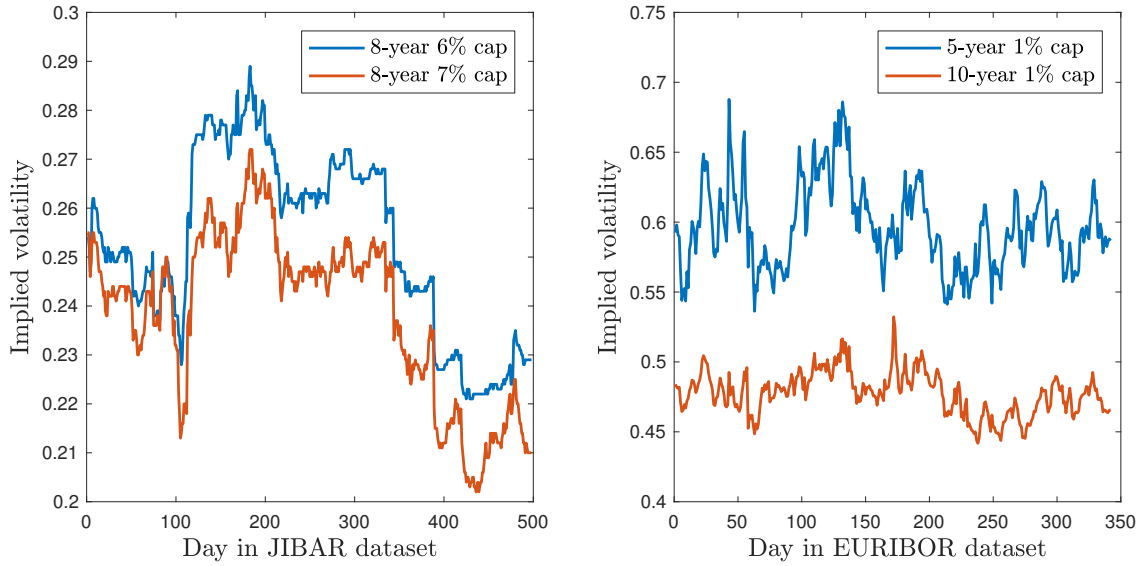


Figure 4.2: Time series of selected (Black) implied volatilities, over the sample periods of either dataset.

Separating the caps into difference caps ensures that each instrument does not overlap with any other (in contrast, the first four caplets appear in all ten caps at a particular strike). For reasons explained in Section 4.2, we exclude the 1- and 2-year difference caps from our study. Note that this separation and exclusion is not possible for at-the-money data, because the at-the-money strike rate depends on the cap term, and so the prices of individual difference caps comprising a longer-term at-the-money cap are not observable. The 6% strike was chosen because it is around the level of 3-month interest rates (the rate underlying the caplets) throughout the sample; while not truly at-the-money, these caps are approximately at-the-money in this informal sense, and so this strike will be given the label ATM. The 7% strike is labelled OTM.

The second dataset is based on EURIBOR-written caps. For 342 trading days (during 2012-2014), we have a bootstrapped OIS and 3-month yield curve, and a term structure of ten implied volatilities for caps struck at 1%. The volatilities are converted, using Black's formula and the two yield curves, into cap prices.⁹ The yield curve is not quite as finely spaced as in the first dataset, and some minor interpolation is required. Again, the caps are arranged into difference caps. The single strike rate of 1% is above the 3-month yield on both curves for the sample period, and is therefore labelled OTM. As this is the lowest fixed strike available, we have no ATM difference caps in our second dataset.

While a lower fixed strike rate among the EURIBOR caps would be desirable, our datasets' bias towards out-the-money (and exclusion of in-the-money) caps is in fact germane to our goal: out-the-money options are more sensitive to volatility, and therefore to possible USV-related hedging error.

Figures 4.1 and 4.2 plot some of the datasets' yield curve and implied-volatility time series, respectively.

⁹See Bianchetti & Carlicchi [2011] for details on the separate use of the OIS-discounting and, in this case, 3-month-forecasting curve. The Johannesburg swap and cap market still functions in a single-curve framework and does not require this.

Hedging Methodology

Following the interest-rate hedging literature, we take ZCBs, implied from the relevant yield curve, as our hedging instruments (for the EURIBOR data, we use the OIS curve to imply ZCB prices).¹⁰ This approach is recommended by its simplicity — other primary market instruments like FRAs and swaps are effectively combinations of several ZCBs — and consistency with the primary market, which in both cases is liquid and easy to trade in.

For our hedging models, we use the G1++ and G2++ models outlined by [Brigo & Mercurio \[2007\]](#). The former is equivalent to the [Hull & White \[1990\]](#) model and the latter its two-factor extension. Both involve Ornstein-Uhlenbeck state variables that, under the risk-neutral measure, revert around a mean of zero:

$$dX_t^{(i)} = -\kappa_i X_t^{(i)} dt + \sigma_i dW_t^{(i)},$$

where κ_i and σ_i are constants and where each $\{W_t^{(i)}\}$ is a (one-dimensional) \mathbb{Q} -Brownian motion (unlike [Example 2](#), the state variables are not forced to have equal mean-reversion rates κ_i). The short rate in the G1++ model is given by

$$r_t = X_t^{(1)} + \varphi(t),$$

where $\varphi(\cdot)$ is a deterministic function of time that allows the model to be calibrated to the initial yield curve. The G2++ model sets the short rate with

$$r_t = X_t^{(1)} + X_t^{(2)} + \varphi(t),$$

where $\varphi(\cdot)$ is again a deterministic function and where the two Brownian motions (involved in the two Ornstein-Uhlenbeck processes) have instantaneous correlation parameter ρ .

For both models, one can find closed-form solutions for ZCBs and ZCB options (and therefore for caps and difference caps) in [Brigo & Mercurio \[2007, Ch.3 and Ch.4\]](#) (the time-dependent function $\varphi(\cdot)$ — absent in [Example 2](#) — causes the bond pricing function to be time inhomogeneous: $P_{tT} = P(t, T, X_t)$). These are used to compute the partial derivatives involved in the following hedging algorithm, which is implemented each day (although, as is made clear below, the hedges are rebalanced each week):

- The model is calibrated to the prevailing yield curve with $\varphi(\cdot)$, and to prevailing difference cap prices with the remaining parameters.¹¹
 - The calibration involves minimising the sum of squared relative pricing errors (as suggested by [Brigo & Mercurio \[2007\]](#)) by varying the model parameters.
 - The five parameters of the G2++ model ($\kappa_1, \kappa_2, \sigma_1, \sigma_2$ and ρ) are calibrated to difference cap prices at each strike separately. This follows [Brigo & Mercurio \[2007\]](#), except that, as mentioned above, we aim to match eight (instead of ten) difference cap prices. The only issue we encounter is a common one that [Brigo & Mercurio \[2007\]](#) discuss: the correlation parameter often has to approach its lower bound of negative one in order to match the curvature often observed in the term structure of cap prices. [Figure 4.3](#) exhibits this by showing a typical calibration.

¹⁰We note an exception in [Gupta & Subrahmanyam \[2005\]](#), who hedge with futures. Because we fix the specific ZCB choice across our difference caps, our approach resembles [Driessen et al. \[2003\]](#) more than [Fan et al. \[2007\]](#) (and [Fan et al. \[2003\]](#)), where the ZCB choice is varied with the instrument under consideration.

¹¹One does not need to compute this function explicitly. In the option formulae, it is only ZCB prices that depend on $\varphi(\cdot)$, and these can be taken directly from the observed yield curve.

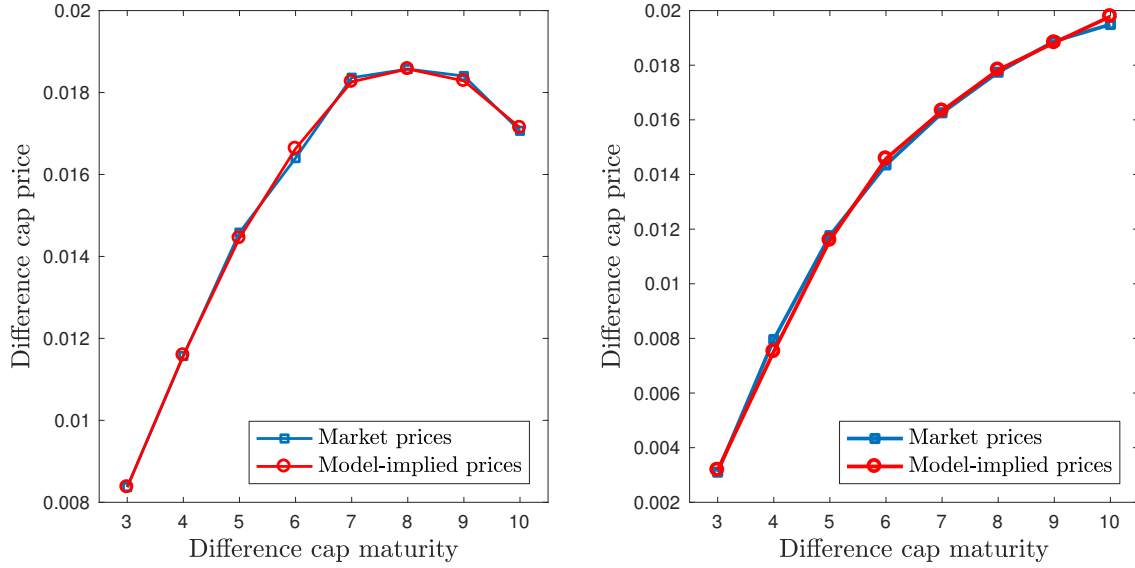


Figure 4.3: Two example calibrations, with prices pertaining to one nominal unit. The left-hand panel shows a term structure of JIBAR-written difference caps struck at 6% — the calibrated parameters are $\kappa_1 = 0.2387$, $\kappa_2 = 0.1076$, $\sigma_1 = 0.0498$, $\sigma_2 = 0.0461$ and $\rho = -0.9076$. The right-hand panel shows EURIBOR difference caps struck at 1%, with calibrated parameters of $\kappa_1 = 0.6649$, $\kappa_2 = 0.1405$, $\sigma_1 = 0.0329$, $\sigma_2 = 0.0235$ and $\rho = -0.99$. Throughout, we constrain the correlation to be no lower than this.

- The two parameters of the G1++ model (κ_1 and σ_1) are insufficient to achieve a reasonable fit to the eight difference cap prices at each strike. The eight are partitioned into the first and second four, and each smaller set can be fitted closely.
- At each calibration one can set $X_0^{(i)} = 0$ without loss of generality. This is because the only effect of non-zero values — given the state variables' dynamics — is a deterministic change to the mean of the future short rate, which can be absorbed into the function $\varphi(\cdot)$.
- Using the calibrated parameters, sensitivities of each difference cap to the state variables are calculated.¹²
 - For each difference cap and for each model, we compute a column vector of sensitivities g , with the i th entry given by

$$[g]_i = \frac{\partial c(x)}{\partial x^{(i)}} \Big|_{x=X_0},$$

where $c(\cdot)$ is the model-implied difference cap price as a function of the state variables.

- ZCBs are engaged so as to match these sensitivities (again, evaluated with the calibrated parameters).
 - For the G1++, there is only one sensitivity and therefore only one ZCB is required (we choose the 3-year bond). For the G2++, two ZCBs are needed to simultaneously hedge two state variable exposures (we choose the 2- and 10-year bonds).

¹²A possible objection to our model choice is that their constant volatilities would not allow changes in volatility to be hedged, to the extent that is possible (in other words, for spanned volatility to be hedged). Our daily calibration, however, allows the volatility parameters, and therefore the hedge ratios, to change as market conditions change. Another possible objection is the low number of factors, but this is only concerning when dealing with swaptions where the correlation between various points of the yield curve is paramount (see Brigo & Mercurio [2007, Ch.4.1] and also the empirical results of the hedging studies in Footnote 10).

- For each model, we define a square sensitivity matrix G with ij th entry given by

$$[G]_{ij} = \frac{\partial P(0, T_j, x)}{\partial x^{(i)}} \Big|_{x=X_0},$$

where $P(\cdot, \cdot, \cdot)$ is the model-implied bond pricing function (for the G1++ model, $T_1 = 3$; for the G2++, $T_1 = 2$ and $T_2 = 10$).

- For each difference cap and for each model, the replicating ZCB positions (in nominal terms) are given by the vector $G^{-1}g$ (this is an application of the so-called *delta-hedging rule* [Shreve, 2004, Ch.4]).^{13,14}
- The remainder — the difference between the difference cap price and the cost of the replicating ZCB(s) — is invested in the cash account.
- The residuals from the hedges put in place *five iterations previously* are recorded — our hedge frequency is thus weekly, but we get as many hedge residuals as we would if we rebalanced daily.¹⁵
 - This requires some interpolation. If, for instance, we are considering the 5-year difference cap, it has now become the 5-year-minus-one-week difference cap, and the 4-year-minus-one-week and 5-year-minus-one-week implied volatilities cannot be directly read off our grid (these are needed to compute the new needed to reprice the 4-year-minus-one-week and 5-year-minus-one-week caps, so that the difference can be taken).¹⁶ We follow the hedging literature in interpolating the required values from the newly prevailing implied volatility surface. Because our strikes are fixed, only the time-to-maturity dimension requires interpolation.¹⁷ Similarly, the 3-year-minus-one-week yield is not directly observable — the newly prevailing yield curve is interpolated, using a cubic spline.
 - If the change in the difference cap exceeds the change in the ZCB-based replicating portfolio, this is recorded as a negative hedge residual — the seller of the difference cap makes a loss in this case.

The above algorithm produces a series of hedge residuals for each difference cap in our two datasets. We produce such a series for both hedging models, as well as a series of unhedged residuals — where no bonds are engaged and the replicating portfolio is simply the cash account — as a point of comparison. This is repeated for data simulated from our benchmark models.

We hedge caps, instead of straddles, as they involve a simpler exposure to the underlying yield curve; they are more easily hedged, which helps us to mitigate the model-dependence and inevitable model-misspecification issue. More generally, the link between the underlying ZCB market and

¹³Definition 2.2.1 referred to a portfolio's nominal ZCB positions as its holdings process. The hedging portfolios in this section conform to the notion of a portfolio given in Definition 2.2.1, in that they consist of ZCB and cash-account holdings. However — as is made clear in the description of the hedging algorithm — these portfolios are not held until the maturity of the option in question; they are held for five days and then compared to the option value, after which the hedging models are reparameterised. The description in terms of a holdings and maturity process, as per Definition 2.2.1, is therefore unnecessary.

¹⁴Note that if the hedging model exhibited an unspanned state variable, G would not be invertible (and the unspanned state variables directions \mathcal{S}^{X_0} from Definition 3.1.1 would give the cokernel of G).

¹⁵This makes our results comparable to the empirical interest-rate hedging literature, where weekly hedging is conducted. It also has the benefit of avoiding deleterious liquidity effects — there is a minor degree of staleness in the JIBAR implied volatilities when viewed at a daily frequency.

¹⁶Note that we therefore do not follow any particular option to its expiry, but continually reset the expiries to their fixed values. This is how the interest-rate hedging literature has made best use of available data. It also ensures that our residual series are stationary — if we followed an instrument to expiry, the properties of the residuals would be expected to change.

¹⁷Usually, we find the interpolation makes no appreciable difference: the observed 5-year implied volatility and the interpolated 5-year-minus-one-week value give effectively the same cap price.

cap market is more straightforward than the ZCB-straddle link; we get a better opportunity of understanding and removing the ZCB-related (spanned) variation among caps, in order to assess whether any unspanned variation remains. In any case, caps are volatility sensitive in their own right, and our benchmark simulations will show them capable of exhibiting USV-induced hedging error.

Benchmark Models

We benchmark our hedging results with hedging based on data simulated from two (very different) models, each of which has a USV and non-USV version.

The first benchmark model (labelled BM1) is from [Bikbov & Chernov \[2009\]](#). Under the real-world measure, BM1 involves a short rate process that satisfies the following coupled dynamics:

$$d \begin{bmatrix} r_t \\ \theta_t \\ v_t \end{bmatrix} = \begin{bmatrix} \kappa_r(\theta_t - r_t) + \kappa_{rv}(\theta_v - v_t) \\ \kappa_\theta(\theta_\theta - \theta_t) \\ \kappa_v(\theta_v - v_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{v_t} & \sigma_{r\theta}\sqrt{(\sigma_\theta)^2 + \beta_\theta v_t} & \sigma_{rv}\sqrt{v_t} \\ \sigma_{\theta r}\sqrt{v_t} & \sqrt{(\sigma_\theta)^2 + \beta_\theta v_t} & \sigma_{\theta v}\sqrt{v_t} \\ 0 & 0 & \sigma_v\sqrt{v_t} \end{bmatrix} dW_t^\mathbb{P},$$

where κ_r , κ_{rv} , κ_θ , κ_v , θ_v , θ_θ , $\sigma_{r\theta}$, σ_θ , σ_{rv} , $\sigma_{\theta r}$, $\sigma_{\theta v}$, σ_v and β_θ are constants and $\{W_t^\mathbb{P}\}$ is a three-dimensional \mathbb{P} -Brownian motion. [Bikbov & Chernov \[2009\]](#) also specify a parametric form for a market-price-of-risk process $\{\zeta_t\}$, which facilitates changing measure, giving the model under \mathbb{Q} . This is based on the *essentially affine* specification of [Duffie \[2002\]](#), which, together with the dynamics above, make BM1 an affine term structure model. Furthermore, in the well-known classification of [Dai & Singleton \[2000\]](#), it is an $\mathbb{A}_1(3)$ model — it has three factors and state variables, one of which drives volatility.

Letting $\{r_t\}$, $\{\theta_t\}$ and $\{v_t\}$ constitute the three-dimensional state process $\{X_t\}$, BM1's (time-homogeneous) bond pricing function is of the form

$$P(\tau, x) = \exp \left(a(\tau) - b_r(\tau)x^{(1)} - b_\theta(\tau)x^{(2)} - b_v(\tau)x^{(3)} \right),$$

where $a(\cdot)$, $b_r(\cdot)$, $b_\theta(\cdot)$ and $b_v(\cdot)$ are deterministic functions that satisfy the ODEs given by [Duffie & Kan \[1996\]](#). If (and only if) five parameter constraints (derived by [Collin-Dufresne & Goldstein \[2002\]](#)) are enforced, there exists a constant β such that

$$b_v(\tau) = \beta b_\theta(\tau)$$

for all $\tau > 0$ (in contrast to the [Fong & Vasiček \[1991\]](#) model considered in Example 1). It follows that the model's unspanned state directions (see Definition 3.1.1) are given by

$$\mathcal{S}^x = \left\{ \begin{bmatrix} 0 & \alpha\beta & -\alpha \end{bmatrix}^\top \mid \alpha \in \mathbb{R} \right\}$$

for all $x \in \mathcal{X} = \mathbb{R}^2 \times \mathbb{R}_+$, which causes the model to exhibit one unspanned state variable (both benchmark models have equal numbers of unspanned factors and state variables; here we only refer to the latter — see the discussion below Definition 3.1.3). The bond covariance function $v(\tau_1, \tau_2, x)$ (Definition 3.2.1) in this model involves many terms, but it is easy to intuit that it is affected by unspanned state movements — these must involve a change in $\{v_t\}$ (and as well as a proportionate change in $\{\theta_t\}$), which features in all of the terms of the bond covariance function, with its coefficients depending on the maturities τ_1 and τ_2 . The set of USV state directions (Definition 3.2.7) is therefore non-trivial, and the model thus exhibits USV.

With and without the USV parameter constraints, [Bikbov & Chernov \[2009\]](#) estimate parameters for this model using quasi-maximum likelihood in conjunction with an extended Kalman filter.¹⁸

We note here that one could easily apply a (linear) \mathcal{S} -transformation — for example, setting

$$\hat{X}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \beta \\ 0 & 0 & -1 \end{bmatrix} X_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \beta \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} r_t \\ \theta_t \\ v_t \end{bmatrix}$$

would satisfy Definition 3.1.14. The third transformed state variable $\{\hat{X}_t^{(3)}\}$ would then be explicitly unspanned and could be removed from the bond pricing function, as per Lemma 3.1.17. We note here that this transformation is not often performed — authors tend refer the untransformed version, as it conforms to the *canonical representation* of $\mathbb{A}_1(3)$ models derived by [Dai & Singleton \[2000\]](#).

The second benchmark model (BM2) is a rational term structure model — based on a specification of a pricing kernel process, as described in Section 1.2 — from [Filipović et al. \[2017\]](#). This makes it much less conventional model than BM1; affine term structure models are the most prevalent in the term structure literature. Their *linear-rational square-root model* gives the pricing kernel as a linear function of the state process:

$$\pi_t = e^{-\alpha t} (\phi + \psi^\top X_t), \quad (4.1)$$

where α and ϕ are constants, ψ is an n -dimensional vector, and where the n -dimensional state process $\{X_t\}$ satisfies

$$dX_t = \kappa(\theta - X_t) dt + \text{diag}\left(\sigma_i \sqrt{X_t^{(i)}}\right) dW_t^\mathbb{P},$$

where each σ_i is a constant, θ is an n -dimensional vector and κ is an $n \times n$ matrix, and where $\{W_t^\mathbb{P}\}$ is an n -dimensional Brownian motion under the real-world measure \mathbb{P} (ensuring an equal number of state variables and factors). Setting $n = 4$ and parameterising κ in a particular way gives rise to the LRSQ(3,1) model of [Filipović et al. \[2017\]](#), which exhibits one USV state variable. In particular, given the bond pricing function (which follows from Equation 1.11) and the form of κ , one finds

$$\mathcal{S}^x = \left\{ \begin{bmatrix} \alpha & 0 & 0 & -\alpha \end{bmatrix}^\top \mid \alpha \in \mathbb{R} \right\}$$

for all $x \in \mathcal{X} = \mathbb{R}_+^4$, so that a linear \mathcal{S} -transform is possible — following Remark 3.1.15, one can define

$$\hat{X}_t = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} X_t,$$

and then describe the model's ZCBs in terms of the first three transformed state variables. [Filipović et al. \[2017\]](#) show, in effect, that the unspanned state directions are also USV state directions, provided that $\sigma_1 \neq \sigma_4$. The model thus exhibits a USV state variable ($\{\hat{X}_t^{(4)}\}$), while not involved

¹⁸They do so both including and excluding option data; we use their option-inclusive estimated parameters.

in the transformed bond pricing function, does feature in the volatility of ZCBs).

Using quasi-maximum likelihood in conjunction with an unscented Kalman filter, [Filipović et al. \[2017\]](#) estimate parameters for their LRSQ(3,1) specification.¹⁹ This provides the USV version of BM2; the non-USV version is attained by freezing the USV state variable at its long-run mean.

Remark 4.1.1. It is noteworthy that in the case of affine term structure models, a USV model is a special case of a non-USV one; in the absence of numerous, knife-edge parameter constraints, the model will by default be fully spanned (and therefore complete, by Theorem 3.1.9). [Filipović et al. \[2017\]](#) show that this is not necessary in the framework of rational term structure models, where a non-USV model can be generalised to exhibit USV. While affine term structure models have many advantageous features in general, they do not provide a natural framework to model USV.

BM1 and BM2 are used to simulate datasets of the same format as our JIBAR data (our larger dataset).²⁰ This involves some numerics. Firstly, both models’ state variables are simulated with an Euler discretisation scheme, using ten time steps per day. It is well known that simulated square-root processes have potential problems at their zero-lower bounds — truncation in the manner of [Lord, Koekkoek & Dijk \[2010\]](#) is used to address this difficulty. The ODE system describing the affine coefficients in BM1’s bond pricing function are numerically solved. In conjunction with the Fourier transform given in [Filipović \[2009, Ch.10\]](#), these ODEs allow the pricing of ZCB options (and therefore caps). For BM2, ZCBs can be priced with a closed-form function. Caps can be priced with a special case of the Fourier swaption formula derived in [Filipović et al. \[2017\]](#), which also relies on affine ODEs.

4.2 Results

We begin, in Section 4.2.1, by interrogating the original evidence of CG and HW, as well as the critique thereof by [Bikbov & Chernov \[2009\]](#). The hedging results are then reported and briefly discussed in Section 4.2.2. Section 4.2.3 contextualises the empirical hedging with hedging results based on the benchmark simulations. This benchmarking shows that the hedging results, in themselves, are inconclusive with regard to USV. The various secondary tests used by FGR and LZ are then scrutinised in Section 4.2.4, which leads to the development and implementation of an improved, conclusive test in Section 4.2.5.

4.2.1 Original USV Evidence

Both CG and HW, as described in Section 4.1.1, regressed a derivative-market quantity against underlying market quantities, supposing that if the latter does not explain the variation in the former, USV factors — which do not directly affect the underlying market — are responsible for the unexplained variance.

In the spirit of HW, we construct a *levels* regression: implied volatility (a key derivative-market quantity) is regressed against the 1-, 3-, 5-, 7- and 9-year yields. A *changes* regression, corresponding to CG’s approach, is defined by regressing weekly changes in implied volatility against weekly changes in the aforementioned yields. The week-long window is rolled forward one

¹⁹This also includes a few parameters that enrich the market-price-of-risk process that is endogenous in the pricing kernel specification. We include this embellishment in our simulations, but, in the interests of clarity, have suppressed it in the description.

²⁰We set the ‘ATM’ strike by first realising the sample path and taking the average of the short-term interest rate. 1% is added to this to determine the OTM strike.

	$\text{mean}(R^2_{\text{levels}})$	$\text{mean}(R^2_{\text{changes}})$	$\% \text{Var}(\text{DPC})$	$\text{mean}(\rho_{\text{res}})$
JIBAR	0.6774	0.1603	90.81	0.8785
EURIBOR	0.4493	0.3736	77.19	0.6697

Table 4.1: Original-evidence metrics — market data.

day at time, so that each day’s data is used. The levels regression is daily. These regressions are estimated using the 4-, 7- and 10-year cap-implied volatilities, at all strike rates.²¹

We have opted to use a set of yields as our underlying market quantities (and will continue to do so below), instead of yield-based principal components (used by HW and in LZ’s secondary tests). We avoid principal components because it would not be clear as to how many should be included — to account for, for instance, 99% of total variance, sometimes only two, but sometimes as many as four, components are needed (when considering our two datasets and simulated paths). The simple set of yields constitutes a reasonable proxy of the yield curve as it affects the cap market, and standardises the number of regressors across datasets. The multicollinearity — the correlation between the various yields and yield changes — has no detrimental effects, as R^2 values (and, in Section 4.2.5, regression residuals) will be the only outputs we consider.

Table 4.1 reports the mean of the regression R^2 s. For the changes regression, which will turn out to be preferable, the proportion of variance of the various regression residuals explained by their dominant principal component (DPC), as well as the mean pairwise correlation coefficient amongst the regression-residual series, are also reported.

The low changes-based regression R^2 , especially from the JIBAR data, is a indication of a disconnect between primary- and derivative-market movements, and therefore of USV. Furthermore, a relatively large proportion of the unexplained derivative-market seems to be chiefly explained by one (presumably USV-related) principal component, giving rise to a strongly cross-correlated series of regression residuals. But, as explained in Section 4.1.1, [Bikbov & Chernov \[2009\]](#) advise caution in this line of interpretation. To benchmark these original-evidence metrics, we simulate five datasets from each version of each benchmark model (described in Section 4.1.2). The identical regression tests are then applied to the simulated data — see Table 4.2 — allowing us to assess whether the R^2 s and other metrics firstly *correctly identify the presence and absence* of USV in *both* benchmark models; and secondly, whether they do so in a manner that is *not detrimentally dependent on the particular realised path*. The intensity of the requisite numerics (particularly the computation of cap prices and the numerical calibration to them) is a significant obstacle — indeed, [Bikbov & Chernov \[2009\]](#) do not apply their benchmarking method to FGR and LZ for this reason — and makes a large number of simulated paths impracticable.²² Five paths from each model version are nevertheless instructive: an adequate sense of the typical behaviour of each model version emerges, as does, importantly, a sense of the variation within each version. This is certainly sufficient to rule out the efficacy of certain metrics, and, we shall see, appears sufficient to vet the metric we develop in Section 4.2.5.

Firstly, we corroborate [Bikbov & Chernov \[2009\]](#) in finding that a levels-based regression is not effective — there is no reliable gap between the two versions of BM2, with the USV version producing R^2 s as large as 0.95. While there is a more successful discrimination between the two versions of BM1, the R^2 s appear to be significantly path dependent.

Secondly, we add to [Bikbov & Chernov \[2009\]](#) in showing the changes-based regression — which

²¹This seems a reasonable way of using the whole surface but also having succinct statistics. Because the volatilities pertain to (overlapping) caps, rather than to separate difference caps, there is no need to take all of the different maturities, which we do below when considering difference caps.

²²While [Bikbov & Chernov \[2009\]](#) do not implement hedges on simulated data due to the computational challenge, they do briefly consider LZ’s findings, which they criticise for the lack of option data in their model estimation.

BM1: non-USV				
Path	$\text{mean}(R^2_{\text{levels}})$	$\text{mean}(R^2_{\text{changes}})$	%Var(DPC)	$\text{mean}(\rho_{\text{res}})$
1	0.9970	0.9735	89.05	0.8669
2	0.9969	0.9917	90.85	0.8834
3	0.9792	0.9647	96.69	0.9617
4	0.9851	0.9812	90.05	0.8832
5	0.9610	0.9492	86.98	0.8481
BM1: USV				
Path	$\text{mean}(R^2_{\text{levels}})$	$\text{mean}(R^2_{\text{changes}})$	%Var(DPC)	$\text{mean}(\rho_{\text{res}})$
1	0.1518	0.0725	99.75	0.9971
2	0.8355	0.1097	99.65	0.9960
3	0.3235	0.2450	99.61	0.9956
4	0.3019	0.1120	99.62	0.9956
5	0.6385	0.1671	99.86	0.9984
BM2: non-USV				
Path	$\text{mean}(R^2_{\text{levels}})$	$\text{mean}(R^2_{\text{changes}})$	%Var(DPC)	$\text{mean}(\rho_{\text{res}})$
1	0.9862	0.8201	96.60	0.9532
2	0.9714	0.9446	95.47	0.9377
3	0.9953	0.9505	92.69	0.8791
4	0.9963	0.9644	80.06	0.8027
5	0.9965	0.9315	84.37	0.7899
BM2: USV				
Path	$\text{mean}(R^2_{\text{levels}})$	$\text{mean}(R^2_{\text{changes}})$	%Var(DPC)	$\text{mean}(\rho_{\text{res}})$
1	0.9098	0.7215	97.00	0.9612
2	0.9385	0.6991	98.68	0.9817
3	0.9542	0.6648	97.39	0.9623
4	0.8819	0.5506	97.43	0.9696
5	0.6494	0.2721	99.21	0.9891

Table 4.2: Original-evidence metrics — simulated data.

they do not consider — to be more successful in identifying the presence of USV. This is especially true for BM1; we replicate CG’s test of their method (their model is affine, similar to BM1). However, we find that this is not perfectly robust across different data-generating processes — the R^2 metric does not discriminate as decidedly when it comes to BM2. The two R^2 ranges do not overlap, but the average difference is not much greater than the variation within each group. This model dependence is interesting in itself and affirms the need to involve multiple models in these experiments. Bringing this to bear on our market data, we find a suggestion of USV presence — the changes-based R^2 s reported in Table 4.1, *in the context of the simulated data*, seem low in the EURIBOR case and very low in the JIBAR case. This, however, is only tentative — BM2 indicates the less than perfect nature of this test — and will need to be supplemented by the hedging results.

Finally, we consider the DPC and correlation metrics, which appear to identify the USV models by attaining values very near to 100% and 1, respectively (especially so for BM1; less reliably for BM2). Unfortunately, though, both non-USV models are capable of producing DPC and correlation metrics as high as 96.5% and 0.96, respectively, making the market-estimated metrics difficult to interpret. There is also a theoretical reason to doubt CG’s interpretation (that large DPC and correlation metrics indicate that the regression residuals are driven by a common USV factor): *multiple* USV factors might be necessary to describe the data. Our simulations — involving models with one USV factor — do not address this possibility, but it is not implausible: a number of studies, such as Filipović et al. [2017], end up favouring models with multiple USV factors.

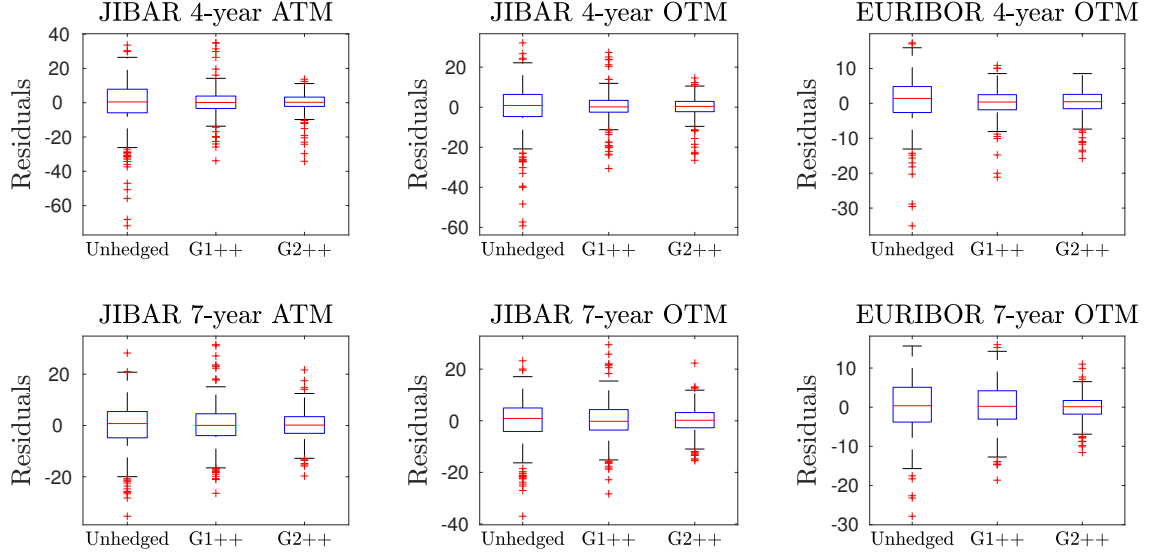


Figure 4.4: Box-and-whisker plots for the hedging residuals of selected difference caps. The residuals are given in basis points, and pertain to difference caps of one nominal unit. An outlier, denoted on the plots in red, is defined as any entry lying outside the first or third quartile by more than one and a half of the interquartile range.

Maturity	JIBAR				EURIBOR	
	G1++		G2++		G1++	G2++
	R^2_{ATM}	R^2_{OTM}	R^2_{ATM}	R^2_{OTM}	R^2_{OTM}	R^2_{OTM}
3	0.829	0.781	0.802	0.746	0.603	0.629
4	0.711	0.651	0.844	0.785	0.692	0.711
5	0.643	0.603	0.860	0.826	0.554	0.834
6	0.402	0.427	0.825	0.788	0.272	0.869
7	0.220	0.235	0.601	0.589	0.292	0.813
8	0.153	0.174	0.476	0.475	0.182	0.744
9	0.141	0.147	0.431	0.418	0.195	0.755
10	0.154	0.189	0.376	0.332	0.148	0.543

Table 4.3: Difference cap hedging R^2 s.

4.2.2 Basic Hedging Results

The hedge residuals series for certain difference caps are summarised in box-and-whisker plots in Figure 4.4 — this provides a slightly more granular view than the hedging R^2 s (defined in Section 4.1.1) reported in Table 4.3. Generally speaking, the difference caps from both datasets are hedged with a reasonable degree of success, comparable to results elsewhere in the literature.²³ In all cases, the hedges manage to reduce at least some variation; in all cases except the 3-year JIBAR caps, the two-factor model manages to reduce more than the one-factor model. Hedging performance is best for mid-dated difference caps; like LZ, our performance deteriorates for longer-term options.

The 1- and 2-year difference caps, which we have excluded, were hedged poorly. This was due to their higher gammas, as well as calibration difficulty.²⁴ These issues manifest only slightly in the 3-year difference caps, which tend to have lower R^2 s than the 4-year maturity. Note that this would not be easily detectable if the caps were not partitioned into difference caps. The partitioning and exclusion helps our study, as the task below becomes to attribute hedge residuals to either model

²³Driessen et al. [2003], when hedging caps on a weekly basis, find R^2 s averaging approximately 0.6; this is similar to our JIBAR-G2++ average (and also to LZ's), and slightly less than our EURIBOR-G2++ one.

²⁴This is both a generally worse model-market fit, as well as occasional explosion of the relative calibration-distance metric — amongst the simulated data, the 1-year OTM cap price occasionally approaches zero.

	$\text{mean}(R_{\text{ATM}}^2)$	$\text{mean}(R_{\text{OTM}}^2)$	$\text{mean}(R^2)$	$\text{mean}(R_{\text{bonds}}^2)$
JIBAR	0.6519	0.6197	0.6358	0.9515
EURIBOR	—	0.7372	0.7372	0.9171

Table 4.4: Hedging results summary — market data.

BM1: non-USV				
Path	$\text{mean}(R_{\text{ATM}}^2)$	$\text{mean}(R_{\text{OTM}}^2)$	$\text{mean}(R^2)$	$\text{mean}(R_{\text{bonds}}^2)$
1	0.9117	0.9050	0.9083	0.9730
2	0.9113	0.9053	0.9083	0.9769
3	0.9015	0.9009	0.9012	0.8831
4	0.9258	0.9079	0.9169	0.9729
5	0.9396	0.9130	0.9263	0.9929
BM1: USV				
Path	$\text{mean}(R_{\text{ATM}}^2)$	$\text{mean}(R_{\text{OTM}}^2)$	$\text{mean}(R^2)$	$\text{mean}(R_{\text{bonds}}^2)$
1	0.8772	0.8338	0.8555	0.9953
2	0.8871	0.8491	0.8681	0.9952
3	0.9380	0.9124	0.9252	0.9950
4	0.8981	0.8561	0.8771	0.9952
5	0.8381	0.7705	0.8043	0.9951
BM2: non-USV				
Path	$\text{mean}(R_{\text{ATM}}^2)$	$\text{mean}(R_{\text{OTM}}^2)$	$\text{mean}(R^2)$	$\text{mean}(R_{\text{bonds}}^2)$
1	0.9050	0.8811	0.8931	0.9714
2	0.9068	0.9208	0.9138	0.9750
3	0.8866	0.8991	0.8929	0.9743
4	0.9093	0.9077	0.9085	0.9662
5	0.9173	0.8979	0.9076	0.9483
BM2: USV				
Path	$\text{mean}(R_{\text{ATM}}^2)$	$\text{mean}(R_{\text{OTM}}^2)$	$\text{mean}(R^2)$	$\text{mean}(R_{\text{bonds}}^2)$
1	0.9039	0.7452	0.8246	0.9656
2	0.8971	0.8528	0.8749	0.9535
3	0.8956	0.8064	0.8510	0.9677
4	0.7223	0.7704	0.7464	0.8386
5	0.8436	0.8589	0.8512	0.9348

Table 4.5: Hedging results summary — simulated data.

misspecification or (possibly present) USV — this is made easier if no significant discretisation- or calibration-related hedging error is present.

4.2.3 Hedging and USV

Our G2++ hedging results are summarised in Table 4.4 (we hereafter consider only the better-performing G2++ model), including the *hedging of bonds*: in addition to the difference caps, we hedge the 1.5-, 3-, 5-, 10- and 15-year ZCBs (with the G2++ model).²⁵ This is done separately for each G2++ calibration, and the overall mean R^2 is reported.

Hedging carried out on simulated data is summarised in Table 4.5 (this includes bond hedging, conducted identically). These results help us address the problematic asymmetry described in Section 4.1.1: it is not clear whether our significantly imperfect hedging is an indication of USV, or is wholly caused by misspecification of our hedging model. Table 4.5 shows how model misspecification manifests, and provides some expectations for hedging results based on USV and non-USV data.

²⁵Note that, as with the caps, the bonds maturities are continually reset so that these maturities are maintained. Recall also that the 10-year ZCB was one of the G2++ hedging instruments — this boundary case is included as a check on our delta and hedge residual calculations.

With regard to difference cap hedging (reported in the first three columns), we firstly observe that the Gaussian hedging model — involving processes quite distinct from those in the benchmark models, and fewer state variables than either — implies hedge ratios that prove reasonably effective, with R^2 s consistently above 0.8. Secondly, the presence of USV detracts from our ability to hedge — but, *at relevant parameter values, the detraction is surprisingly small* (with an R^2 difference, averaging over both models, of 0.06). FGR’s effective hedging is therefore not as strongly indicative of the absence of USV as one might think at face value — their hedging effectiveness is in fact similar to that found for the USV versions of both benchmark models. Thirdly, The USV R^2 s, though not much smaller on average, are *much more varied* than the almost constant non-USV R^2 s — the USV factor’s activity and effect on hedging effectiveness is rather dependant on the particular realised path. Finally, and more generally, we conclude that the hedging results on their own are not decisive — given that USV, if present, would only have a small impact on the observed hedging R^2 s in Table 4.4, these metrics have no clear USV implication. Our attention therefore turns, in the next subsection, to the secondary tests described in Section 4.1.1.

With regard to bond hedging, recall from Section 4.1.1 that LZ’s primary USV evidence is the contrast between their cap and bond hedging effectiveness. A very similar contrast is displayed in Table 4.4.

One might expect the effectiveness of bond hedging to be impervious to the presence of USV. However, typically and indeed in this study, at least some of the model parameters are determined by calibration to cross-sectional option prices, which are affected by USV factors if present. The bond hedge ratios are thus susceptible to this irrelevant, random influence. This is borne out by BM2. When applying the USV parameter restrictions to BM1, Bikbov & Chernov [2009] estimate a significantly lower mean-reversion level for the volatility process $\{v_t\}$ and this influence appears to outweigh the effect of the turbulent calibrations.²⁶ All of this suggests that the contrast to bond hedging effectiveness is not particularly informative. While a bond-hedging R^2 of 98%, say, might appear to testify to an excellent model that can only fail in the presence of USV, bonds on the same yield curve are in fact easy instruments to cross hedge — using simple duration-based hedge ratios to the same bonds (both those hedging and being hedged), we in fact outperform the G2++ R^2 on the JIBAR data, and nearly match it on the EURIBOR data. Thus, effective bond hedging is not necessarily an indication of a model having captured term structure dynamics particularly well.

4.2.4 Secondary Tests

Based on the the secondary tests used by FGR and LZ (detailed in Section 4.1.2), we define four regressions and give each a simple label:

- LZa: we regress *the cross-sectional sum* of G2++ hedge residuals (expressed in basis points per unit-nominal difference caps) against the corresponding changes in the set of yields considered in Section 4.2.1;
- LZb: we add changes in the 10-year OTM implied volatility (which exists in every dataset) to the set of regressors in LZa;
- FGRa: the same hedge residuals of LZa are regressed against changes in implied volatility for all the maturities and strikes hedged. Except for the EURIBOR data, there are sixteen such volatilities (FGR used all of their twenty four); and

²⁶This mean-reversion level parameter is θ_v in our notation; \bar{v} in theirs.

	BM1: non-USV				
Path	R^2_{LZa}	R^2_{LZb}	R^2_{FGRa}	R^2_{FGRb}	%Var(DPC)
1	0.9324	0.9466	0.9536	0.3023	86.01
2	0.9744	0.9765	0.9775	0.5991	65.98
3	0.9135	0.9574	0.9727	0.4839	75.89
4	0.9423	0.9704	0.9662	0.5927	74.24
5	0.9443	0.9699	0.9773	0.6624	75.56
	BM1: USV				
Path	R^2_{LZa}	R^2_{LZb}	R^2_{FGRa}	R^2_{FGRb}	%Var(DPC)
1	0.0481	0.9868	0.9793	0.7639	89.96
2	0.0098	0.9745	0.9768	0.7727	90.33
3	0.0272	0.9829	0.9777	0.5275	81.82
4	0.0813	0.9886	0.9692	0.6143	92.87
5	0.0101	0.9909	0.9925	0.7690	92.99
	BM2: non-USV				
Path	R^2_{LZa}	R^2_{LZb}	R^2_{FGRa}	R^2_{FGRb}	%Var(DPC)
1	0.7943	0.8164	0.8598	0.1514	72.67
2	0.8380	0.8409	0.8162	0.0007	90.42
3	0.8623	0.8922	0.9087	0.1966	91.68
4	0.8585	0.8926	0.7422	0.0150	88.54
5	0.7993	0.8596	0.5777	0.0971	89.96
	BM2: USV				
Path	R^2_{LZa}	R^2_{LZb}	R^2_{FGRa}	R^2_{FGRb}	%Var(DPC)
1	0.8297	0.9668	0.8334	0.5196	81.84
2	0.8362	0.9482	0.8613	0.3082	75.73
3	0.7932	0.9151	0.7382	0.4004	82.19
4	0.6138	0.8234	0.5509	0.1734	94.83
5	0.6532	0.8907	0.7651	0.2105	87.78

Table 4.6: Secondary test results — simulated data.

- FGRb: the residuals are regressed against changes in the OTM 10-year implied volatility. This more parsimonious regression is added to compare to the highly parameterised FGRa.

We also calculate the proportion of variance in the series of difference cap hedge residuals explained by their DPC (this involves neither a cross-sectional sum nor a regression), which both FGR and LZ adduce as secondary evidence. This and the R^2 s for the above regressions are reported in Table 4.6, which leads to several findings.

For BM1, the LZa R^2 is extremely informative in the way intended by LZ: with and only with USV can the underlying yield changes market fail to explain the observed hedging errors. However, this turns out to be significantly dependent on the particular model and realised path — for BM2, R^2_{LZa} tends to be slightly lower for the USV data, but not significantly or consistently enough to make it a powerful general test. To briefly recapitulate the discussion in Section 4.2.1, we are seeking a metric that identifies the presence *and* absence of USV, in a way that does not depend greatly on the particular path; in other words, an effective test metric is one where the USV values are clearly separate from the non-USV ones, with the separation being greater than the variation within a particular model version. Furthermore, this must be robust to a change in the underlying model (e.g., from BM1 to BM2). Then, when the metric is calculated on market data (the precise data-generating nature of which is unknown), the result can be interpreted unequivocally.

The next metric of interest is the *difference* between the LZa and LZb R^2 s; if the addition of derivative-market information (for which implied volatility is a proxy) significantly increases explanatory power of hedging residuals, then the derivative market contains (unspanned) information beyond that contained in the underlying yields originally used in LZa. When considering this R^2

Cause (increase in)	Effects		
	Underlying market	Derivative market	Hedge residuals
Spanned factor	yes	yes	yes
USV factor	no	yes, positive	yes, negative

Table 4.7: Cause-and-effect schematic.

difference, BM1 remains well handled and BM2’s handling is improved: $R_{LZb}^2 - R_{LZa}^2$ is smaller than 0.06 in non-USV paths, and larger than 0.11 when USV is present, giving the metric some practical validity. However, this separation between the USV and non-USV cases is not satisfactorily large, especially considering the path dependence within each model version. We seek to improve this in the next subsection.

The FG Ra regression is likely over-parameterised, resulting in large R^2 s that do not reveal the presence of USV in either benchmark model. FG Rb performs reasonably on BM2 — when USV is contributing to the hedging residuals, more hedge-residual variation can be explained by implied-volatility (i.e., derivative-market) variation. However, the USV and non-USV metric sets still overlap, and the metric fails to discriminate the versions of BM1.

Our simulation context is favourable for the principal component metric, in that the complicating possibility of multiple USV factors (mentioned in Section 4.1) can be ignored.²⁷ Even with this benefit, it fails to distinguish the USV and non-USV generated data.

4.2.5 The Two-tier Test

The goal of the regressions in the previous subsection is to determine, by the observed R^2 s, whether hedge residuals are driven simply by movements in the underlying yields, or whether there is a significant contribution from USV factors (as discussed in Footnote 2, this terminology of factors is a pragmatic shorthand). A brief discussion of the fundamental cause-and-effect relationships will assist in constructing a more effective regression. In particular, whether a certain relationship is monotonic is of interest — this is a necessary criterion for a linear regression to meaningfully capture a relationship.

Movements in a USV factor (if one exists) affect the derivative market but do not affect the underlying yields; because the latter drive our hedge instruments, a hedge residual is induced by a USV factor movement and would be reflected in the cross-sectional sum used in LZa and LZb. Furthermore, a USV factor *increase* will have a *positive* effect on conventional, positive-vega derivatives such as caps; by our convention, the induced hedge residual is *negative*. Changes to spanned factors (i.e., changes in the yield curve) affect both the underlying and derivative markets, in a way that depends on the particular change. Because the hedging model is inevitably misspecified to a degree, a hedge residual also results. This cause-and-effect discussion is summarised in Table 4.7. The question of interest, from an empirical-USV perspective, is whether the events of the type in the second row occur with any significance.

Given this framework, it can be seen that regressions involving derivative-market proxies such as implied volatility — the FGR regressions — are not apt to identify USV: changes in any type of factor cause movements in the derivative market and in implied volatility, and also cause hedge residuals. A relationship between hedge residuals and implied volatility, with a strength unknown a priori, will likely emerge whether USV is present or not.

Because the LZa regression takes underlying-yield information as its explanatory variables, it

²⁷This complication is illustrated in how FGR and LZ, both based on low variation ascribed to the DPC, reach opposing conclusions: FGR claim no USV factors are necessary to explain their data, while LZ conclude multiple such factors are needed.

	BM1: non-USV		BM1: USV	
Path	$R^2_{\text{second-tier}}$	$\beta_{\text{second-tier}}$	$R^2_{\text{second-tier}}$	$\beta_{\text{second-tier}}$
1	0.0051	-0.7316	0.8487	-40.3763****
2	0.0006	-0.1583	0.7996	-38.5336****
3	0.0136	-1.1056**	0.5839	-22.7727****
4	0.0058	-0.7096	0.7537	-35.2822****
5	0.0154	-1.1399**	0.7143	-35.7308****
	BM2: non-USV		BM2: USV	
Path	$R^2_{\text{second-tier}}$	$\beta_{\text{second-tier}}$	$R^2_{\text{second-tier}}$	$\beta_{\text{second-tier}}$
1	0.0091	-2.5838*	0.2646	-12.4554****
2	0.0005	-0.2942	0.2062	-14.6480****
3	0.0008	-0.3567	0.2384	-10.4527****
4	0.0030	-0.6025	0.2469	-11.6903****
5	0.0233	-1.5650***	0.4277	-19.4378****

Table 4.8: Two-tier test results — simulated data.

can only explain residuals that result from spanned factor movements (corresponding to the first row in Table 4.7). If a USV factor is present, its induced residuals cannot be explained by this regression. Instead of expanding the set of regressors (from LZa to LZb) and observing whether these USV-related residuals can be explained, we propose the following *two-tier* alternative:

1. Perform the LZa regression.
 - Hedge residuals due to imperfectly hedged spanned factor movements are identified to the extent facilitated by the linear regression. There is no natural monotonicity here — we do not know whether an increase in the 5-year yield, for instance, would induce a hedging profit or loss — but, with the benefit of hindsight, the regression will identify the available correlations.
 - Note that these yield-curve regressors cannot, in principle, explain the potential residuals from USV factors.
2. Regress the residuals from the above regression against changes in the 10-year OTM implied volatility (the derivative-market proxy).
 - The residuals from the initial LZa regression include the residuals resulting from USV, if in fact present to any significant degree.
 - If USV is present, one would expect this second regression to find a *decreasing* relationship with some explanatory power. This is because an increase in an USV factor increases the price of a caps — increasing implied volatilities — without a corresponding change in the hedge instruments, resulting in a negative hedge residual. The test metrics are therefore the R^2 of this second-tier regression and the slope coefficient it finds.
 - If USV has no significant presence, we expect the step 2 regression to have low explanatory power — there is no reason to think the dependent variable will be meaningfully related to a volatility proxy.

The crucial step is to test this proposal on simulated data and ensure that it can discriminate in the intended way — see Table 4.8. Note that *, **, *** and **** denote coefficient significance at the 5%, 1%, 0.1% and 0.01% level, respectively; that is, the p -value from supposing that the regression's slope coefficient is zero is at least as small as the significance level indicated.

	R_{LZa}^2	R_{LZb}^2	R_{FGRa}^2	R_{FGRb}^2	%Var(DPC)
JIBAR	0.2471	0.8922	0.7203	0.2836	30.91
EURIBOR	0.4735	0.7045	0.3986	0.0505	35.77

Table 4.9: Secondary test results — market data.

JIBAR		EURIBOR	
$R_{\text{second-tier}}^2$	$\beta_{\text{second-tier}}$	$R_{\text{second-tier}}^2$	$\beta_{\text{second-tier}}$
0.6926	-41.6902****	0.3094	-3.5109****

Table 4.10: Two-tier test results — market data.

The results are encouraging: for both benchmark models, the R^2 of the newly proposed test is *significantly and consistently higher* when USV is present in the data. BM1 remains the more amenable of the benchmark models, but BM2 also conforms to the ex-ante expectations: although there is some path-dependence in the R^2 s, the lowest value in the USV case is almost an order of magnitude larger than the highest in the non-USV case. Additionally, as expected, a negative and statistically significant regression coefficient is found in the USV case (in every case, at the 0.01% level). While the non-USV cases happen to find a negative relationship, it is consistently less pronounced and seldom significant.

Tables 4.9 and 4.10 report, respectively, the aforementioned secondary test metrics and the two-tier test results for our datasets.

Viewing the two-tier test results (in Table 4.10) in the context of the benchmark simulations (in Table 4.8), we reach a decisive conclusion: the R^2 s from both datasets, being significantly larger than the lowest R^2 generated by a USV benchmark model, provide a strong indication of USV. The highly significant, negative coefficients align with this view.

Given the degree of separation seen in Table 4.8, the small number of simulated paths do not appear an obstacle to this conclusion — the non-USV simulations make it seem extremely unlikely that the significant, negative second-tier relationships reported in Table 4.10 (for both datasets) are caused solely by imperfectly hedged yield movements.

4.3 Conclusion

Our hedging of difference caps simulated from benchmark models shows that USV (in the degree implied by the empirically based parameters) does not in fact have a great impact on one’s ability to hedge. Observed hedging effectiveness is therefore not especially informative about the possible presence of USV, resolving the apparent conflict between the results of FGR and LZ. More generally, the results of Section 4.2.3 show that it might be reasonable for a hedger of interest-rate derivatives to overlook unspanned sources of variation — on average, hedging effectiveness deteriorates only marginally in the face of USV, although this effectiveness becomes more variable.

This does not imply that USV-generated data is virtually the same as non-USV-generated data. They are quite different, in important conceptual senses described in Section 1.2. Indeed, in Section 4.2.5 we show that the manifestations of USV can be clearly identified. Despite our finding that USV cross-sectional relationships are *relatively* minor and have rather minors effects on hedging effectiveness, USV has clear discernible impacts on time-series behaviour.

Our proposed two-tier test — based on the fundamental relationships involved in hedging residuals, and vetted by our benchmark model simulations — gives a strong indication of USV in both datasets. This is in alignment with the tentative USV indication from CG’s original regression (which our benchmark simulations show to be unsatisfactorily robust) and also with the

lack of indications throughout our results of the opposite, non-USV, conclusion (in particular, the difference between the LZb and LZa R^2 s — reported in Table 4.9 — is large for both our datasets).

Strictly speaking, the conclusions are limited to our two particular datasets, but there are no reasons to think that most other fixed-income markets are structurally different. Our conclusions can therefore be tentatively taken to apply generally. USV models are necessary to plausibly model the interest-rate markets.

Although both datasets yield the same conclusion, the magnitudes of the regression coefficients reported in Table 4.10 are quite different. Although always highly significant, the coefficients from our two USV benchmark models also differ in size. Future research could interrogate the role of the low-interest regime in the JIBAR-EURIBOR coefficient discrepancy, and how this would manifest in hedging and risk management.

Chapter 5

Bivariate USV Models

We begin by further specialising the framework of Section 2.1. The chapter is concerned with bivariate TSMs, which motivates the first specialising assumption.

Assumption 5.0.1. Recalling that n denotes the dimension of the state process $\{X_t\}$ and d the dimension of the factor process $\{W_t\}$, we assume that $n = d = 2$.

Bivariate models are thus taken as TSMs involving two factors *and* two state variables. This notion therefore excludes the two-factor instance of the Trolle & Schwartz [2009] model discussed in Section 3.4.2.

We will sacrifice a slight degree of generality by insisting that TSMs can be *cast in terms of their short rate* (the short-rate process $\{r_t\}$ — which gives the instantaneous return available in the cash account — is defined in Equation 2.2).

Assumption 5.0.2. The first state variable $\{X_t^{(1)}\}$ is given by the short-rate process $\{r_t\}$.

The short rate is often taken as a state variable in the first place (this is discussed briefly in Section 1.2). Where it is not, we are assuming that there exists a state process transformation (Definition 2.2.7) that gives the short rate as one of the transformed state variables.¹ An example of such a transformation can be seen in Appendix D.1.

Letting $\{u_t\}$ denote the second state variable, from Assumption 2.1.5 we have that

$$d \begin{bmatrix} r_t \\ u_t \end{bmatrix} = \mu(t, r_t, u_t) dt + \sigma(t, r_t, u_t) dW_t. \quad (5.1)$$

Assumption 5.0.3. Introducing notation to denote the sub-processes involved in Equation 5.1, we assume that

$$d \begin{bmatrix} r_t \\ u_t \end{bmatrix} = \begin{bmatrix} \mu^{\mathbb{Q},r}(t, r_t, u_t) \\ \mu^{\mathbb{Q},u}(t, r_t, u_t) \end{bmatrix} dt + \begin{bmatrix} \sigma^r(t, r_t, u_t) & 0 \\ \sigma^{u,1}(t, r_t, u_t) & \sigma^{u,2}(t, r_t, u_t) \end{bmatrix} dW_t.$$

The assumption of a lower-triangular diffusion matrix is convenient but sacrifices no generality.²

With Assumptions 5.0.1 and 5.0.2, our specialised framework captures all *bivariate, short-rate models* encompassed by Section 2.1. Our interest now is in the circumstances under which these

¹In other words, we assume that the short rate can be accompanied by another function of the original state variables such that the transformation is invertible.

²Conceptually, this is because it is only the variance matrix — the volatility matrix multiplied by its transpose — that affects that state variables' distribution, and any variance matrix can be attained by some lower triangular volatility. Indeed, this assumption amounts to pre-applying the orthonormal factor transformation described in the first subsection of Section 3.3.3 (see Remark 3.3.5). See Appendix D.2 for an example of this transformation.

models exhibit USV. For the model to exhibit a USV factor as per Definition 3.2.5, two conditions must hold. Firstly, there must exist a dimension of unspanned factor directions (in a non-null part of the state space). If we assume that an \mathcal{S} -transformation (Definition 3.1.14) has been applied, any unspanned directions correspond to the second state variable. Secondly, movements in this unspanned direction — i.e., movements to the explicitly unspanned second state variable — must affect bond covariances in some way. These two conditions motivate the following definition. The way that the first is reflected is based on the idea of a *reduced bond pricing function* from Lemma 3.1.17 — the unspanned state variable can be removed from the bond pricing function so that $P_{tT} = P(t, T, r_t)$.

Definition 5.0.4. The TSM in question is a $(1,1)$ model if and only if ZCB prices can be given by a reduced pricing function and $\sigma^r(\cdot, \cdot, \cdot)$ (from Assumption 5.0.3) is a non-trivial function of its third argument (i.e., $\sigma^r(t, r, u)$ cannot be given as a function of only t and r).

In other words, a bivariate short-rate model is a $(1,1)$ model when the bond pricing function *can* be reduced (so that $\{u_t\}$ is unspanned) *and* the short-rate volatility function depends on $\{u_t\}$ and thus *cannot* be reduced.³ Note also that Definition 5.0.4 excludes a constant short rate, by indirectly ensuring that the short rate volatility is non-zero (lest it be a non-trivial argument of u).

The process $\{\sigma^{u,2}(t, r_t, u_t)\}$ (introduced in Assumption 5.0.3) would usually be almost-surely non-zero, so that the model's diffusion matrix is full rank.⁴ Assuming this, one can easily verify that a TSM satisfying Assumptions 5.0.1, 5.0.2 and 5.0.3 and Definition 5.0.4 exhibits one USV factor. The only bivariate USV models excluded by our specialised framework are those for which an \mathcal{S} -transformation (that gives the spanned component as the short rate) does not exist. Such models, despite being bivariate and exhibiting USV, can be excluded by Definition 5.0.4, as their bivariate nature cannot be explicitly represented by pair of state variables of which one is unspanned (i.e., represented in a $(1,1)$ fashion).

Without assuming that Definition 5.0.4 holds, the bond pricing function takes both state variables, as well as calendar and maturity time; writing the dependence on either state variable separately, we have $P_{tT} = P(t, T, r_t, v_t)$ (as the relevant special case of Equation 2.1). It is well known that the bond pricing function must satisfy a certain PDE, which we refer to as the *term structure equation*.⁵ In the case of a $(1,1)$ model, where the bond pricing function is reduced, the term structure equation is given by

$$P_t(t, T, r) + P_r(t, T, r)\mu^{\mathbb{Q},r}(t, r, u) + \frac{1}{2}P_{rr}(t, T, r)\sigma^r(t, r, u)^2 = rP(t, T, r), \quad (5.2)$$

which must hold for all $T > 0$, $t \in [0, \min(T, S)]$ and $[r \ u]^\top \in \mathcal{X}$. The subscripts denote partial derivatives with respect to the various arguments of the bond pricing function — this notation will be used throughout the chapter.⁶

With the dependence on the two state variables written separately, the bond pricing function

³In contrast, Lemma 3.3.10 does reduce the diffusion function of the spanned state variables, based on the assumption of a lack of USV.

⁴If it were not, the model would exhibit a USV state variable but not a USV factor, which, as is discussed in Remark 3.2.12, would not result in a meaningful exhibition of USV.

⁵Like the function $q(\cdot, \cdot)$ in Section 3.3.2, $P(\cdot, \cdot, \cdot, \cdot)$ is the Feynman-Kac stochastic representation of a PDE solution. See Björk [2004, Ch.23] for a derivation of Equation 5.2.

⁶Note that $P_t(\cdot, \cdot, \cdot)$, for example, denotes the partial derivative of $P(\cdot, \cdot, \cdot)$ with respect to its first argument. This, for example, implies that

$$P_t(t, t, r) = \frac{\partial P(t, T, r)}{\partial t} \Big|_{(t, T, r) = (t, t, r)} \neq \frac{\partial}{\partial t} P(t, t, r).$$

is $\mathcal{C}^{1,1,2,2}$ by Assumption 2.1.2. We assume an increased degree of continuous differentiability in the maturity-time argument, the necessity of which is discussed in Remark 5.1.2 below.

Assumption 5.0.5. $P(\cdot, \cdot, \cdot, \cdot)$ is a $\mathcal{C}^{1,2,2,2}$ function.

The chapter proceeds as follows. Section 5.1 presents the chapter's central theoretical result. In Section 5.2, a specific bivariate USV model is proposed and analysed in detail. Section 5.3 briefly outlines some extensions.

5.1 (1,1) Models

Lemma 5.1.1. For a (1,1) model, we have

$$\lim_{T \downarrow t} \frac{P_{rr}(t, T, r)}{P_r(t, T, r)} = 0.$$

Remark 5.1.2. The intuition of Lemma 5.1.1 is as follows. As maturity time T is brought near to calendar time t , ZCBs approach their deterministic maturity value of one. There is then no sensitivity to the short rate ($P_r(t, t, r) = P_{rr}(t, t, r) = 0$), making the above limit an indeterminate form. As the time to maturity shortens, a ZCB price becomes well approximated by the short rate level (as the short rate is precisely the return offered by a short-term ZCB). In particular, the second-order derivative implied by this approximation approaches zero faster than the first-order derivative, giving the limit a value of zero. Second-order differentiability in maturity time (Assumption 5.0.5) is needed to ensure that approximation validly leads to the conclusion, as the limit is taken.⁷

Note also that this result, although stated for (1,1) models, applies to any TSM with a bond pricing function that is differentiable in the short rate and that satisfies the differentiability assumptions of the chapter.

Proof. Consider a second-order Taylor expansion of the bond pricing function $P(t, T, r)$ in its second argument T , around the point $T = t$ (Assumption 5.0.5 gives the necessary twice differentiability). Using the subscript-partial-derivative notation introduced above, one attains

$$P(t, T, r) = P(t, t, r) + P_T(t, t, r)(T - t) + \frac{1}{2}P_{TT}(t, \bar{T}, r)(T - t)^2 \quad (5.3)$$

for some $\bar{T} \in (t, T)$. Appendix C.1 shows that $P_T(t, t, r) = -r$, and discusses this alternative to the standard link between the short rate and bond pricing function given in Equation 2.2. One can then proceed with

$$P(t, T, r) = 1 - r(T - t) + \frac{1}{2}P_{TT}(t, \bar{T}, r)(T - t)^2.$$

This leads to the following representation of the bond price's (assumed existent) partial derivative with respect to the short rate:

$$P_r(t, T, r) = -(T - t) + \frac{(T - t)^2}{2} \frac{\partial}{\partial r} P_{TT}(t, \bar{T}, r). \quad (5.4)$$

⁷Without this assumption, instead of Equation 5.3, one gets

$$P(t, T, r) = P(t, t, r) + P_T(t, t, r)(T - t) + h(t, T, r)(T - t),$$

with the assurance that $\lim_{T \downarrow t} h(t, T, r) = 0$, but no assurance about the limiting behaviour of $\frac{\partial h(t, T, r)}{\partial r}$ or $\frac{\partial^2 h(t, T, r)}{\partial r^2}$.

Differentiating again, one gets

$$P_{rr}(t, T, r) = \frac{(T-t)^2}{2} \frac{\partial^2}{\partial r^2} P_{TT}(t, \bar{T}, r). \quad (5.5)$$

Assuming that $P_r(t, T, r) \neq 0$, we then have

$$\frac{P_{rr}(t, T, r)}{P_r(t, T, r)} = \frac{(T-t) \frac{\partial^2}{\partial r^2} P_{TT}(t, \bar{T}, r)}{(T-t) \frac{\partial}{\partial r} (P_{TT}(t, \bar{T}, r)) - 2}. \quad (5.6)$$

Appendix C.2 establishes that there necessarily exists some positive δ such that $P_r(t, T, r) \neq 0$ for all $T \in (t, t + \delta)$. Equation 5.6 is therefore well-defined when T is sufficiently close to t , and can be used to assess the limit given in the lemma. As T approaches t , the numerator approaches a value of zero, and the denominator a value of -2 (note that, by Assumption 2.1.2 or 5.0.5, Equations 5.4 and 5.5 must have finite values). The limit thus has a value of zero, proving the lemma. \square

Our main result thoroughly characterises (1,1) models.

Theorem 5.1.3. Assuming that the first-order initial value problem in Equation 5.13 below admits a solution, the following are equivalent:

- (A) The model is a (1,1) model, as per Definition 5.0.4;
- (B) $P_{tT} = g(t, T) - f(t, T)r_t$;
- (C) $\mu^{\mathbb{Q}, r}(t, r, u) = r^2 + r\alpha_1(t) + \alpha_2(t)$ for deterministic functions $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$.

Remark 5.1.4. The assumption of a solution to Equation 5.13 — which is required for (C) to imply (B) — is not restrictive in practice, because the existence of a solution depends on whether $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ are well-behaved, and one would not be greatly interested in a specification involving irregular $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$. Furthermore, in Corollary 5.1.5 below, we show that Equation 5.13 does indeed admit a solution in our primary case of interest.

Proof. Noting that (B) trivially implies (A), we begin by proving that (A) implies (B). We follow Collin-Dufresne & Goldstein [2002, Appx.B], where it is noted that the term structure equation — which, under (A), is given by Equation 5.2 — applied at two maturity times T_1 and T_2 (both greater than t) can be written

$$\begin{bmatrix} \frac{1}{2}P_{rr}(t, T_1, r) & P_r(t, T_1, r) \\ \frac{1}{2}P_{rr}(t, T_2, r) & P_r(t, T_2, r) \end{bmatrix} \begin{bmatrix} (\sigma^r(t, r, u))^2 \\ \mu^{\mathbb{Q}, r}(t, r, u) \end{bmatrix} = \begin{bmatrix} rP(t, T_1, r) - P_t(t, T_1, r) \\ rP(t, T_2, r) - P_t(t, T_2, r) \end{bmatrix}.$$

Collin-Dufresne & Goldstein [2002] point out that if the matrix on the far left were invertible, its inverse could be multiplied to either side and, from viewing the right-hand side, would show that $\sigma^r(t, r, u)$ is not a function of u . Because this is precluded by Definition 5.0.4, the far left matrix is not invertible, and its determinant is zero. We therefore have

$$P_{rr}(t, T_1, r)P_r(t, T_2, r) - P_r(t, T_1, r)P_{rr}(t, T_2, r) = 0. \quad (5.7)$$

Equation 5.7 must hold for any maturities $T_1 > t$ and $T_2 > t$. If we consider taking T_2 increasingly close to t , Lemma 5.1.1 states that the ratio

$$\frac{P_{rr}(t, T_2, r)}{P_r(t, T_2, r)}$$

is well-defined and approaches zero. For small values of T_2 then, Equation 5.7 can be written

$$P_{rr}(t, T_1, r) - P_r(t, T_1, r) \frac{P_{rr}(t, T_2, r)}{P_r(t, T_2, r)} = 0.$$

and because the second term becomes arbitrarily close to zero, we must have that

$$P_{rr}(t, T_1, r) = 0$$

for all $t < T_1$ and r . Integrating twice with respect to r , we see that the bond pricing function (at any maturity time T_2) is affine in the short rate, completing this section of the proof. Note that the constants of integration may depend on the two other arguments of $P(\cdot, \cdot, \cdot)$, giving the form written in (B).

To prove that (B) implies (C), the pricing function assumed in (B) allows the term structure equation (given, under (B), by Equation 5.2) to simplify to

$$g_t(t, T) - f_t(t, T)r - f(t, T)\mu^{\mathbb{Q}, r}(t, r, u) + 0 = r(g(t, T) - f(t, T)r). \quad (5.8)$$

Differentiating twice with respect to r yields

$$-f(t, T) \frac{\partial^2 \mu^{\mathbb{Q}, r}(t, r, u)}{\partial r^2} = -2f(t, T),$$

or, rewriting slightly,

$$-f(t, T) \left(\frac{\partial^2 \mu^{\mathbb{Q}, r}(t, r, u)}{\partial r^2} - 2 \right) = 0.$$

It follows that⁸

$$\frac{\partial^2 \mu^{\mathbb{Q}, r}(t, r, u)}{\partial r^2} = 2.$$

Integrating twice, and allowing the constants to depend on the other arguments t and u , we have

$$\mu^{\mathbb{Q}, r}(t, r, u) = r^2 + \alpha_1(t, u)r + \alpha_2(t, u). \quad (5.9)$$

Substituting this into Equation 5.8, and rearranging terms, we have

$$g_t(t, T) - f(t, T)\alpha_2(t, u) - r \left(f_t(t, T) + f(t, T)\alpha_1(t, u) + g(t, T) \right) = 0.$$

Under (B), this equation must hold for all r . Therefore, the first two terms (which are independent of r) and the coefficient of $-r$ (in the large brackets) must separately be zero. This implies that

$$\alpha_2(t, u) = \frac{g_t(t, T)}{f(t, T)} \quad \text{and} \quad \alpha_1(t, u) = \frac{-g(t, T) - f_t(t, T)}{f(t, T)},$$

which shows that $\alpha_1(\cdot, \cdot)$ and $\alpha_2(\cdot, \cdot)$ do not depend on their second argument u . Together with Equation 5.9, this proves that (B) implies (C).

Finally, to demonstrate that (C) implies (B), we use the standard method employed famously by Duffie & Kan [1996]: we postulate a form for the bond pricing function, and then verify that,

⁸Note that the first term cannot equal zero for all maturities — Appendix C.2 establishes that for a sufficiently small maturity time T^* , there is non-zero partial derivative with respect to the short rate: $P_r(t, T^*, r) = -f(t, T^*) \neq 0$.

along with the drift from (C), the relevant term structure equation is satisfied (as well as verifying that the pricing function has the correct terminal condition). We postulate that the bond pricing function is affine in the short rate, precisely as in condition (B), with, for each $T > 0$, $f(\cdot, T)$ satisfying

$$f_{tt}(t, T) + f_t(t, T)\alpha_1(t) + f(t, T)\left(\frac{\partial\alpha_1(t)}{\partial t} + \alpha_2(t)\right) = 0 \quad (5.10)$$

for $t \in [0, T]$, along with the terminal conditions $f(T, T) = 0$ and $f_t(T, T) = -1$ (note that $\alpha_1(\cdot)$ is differentiable by Assumption 2.1.5). This linear, second-order ODE is reduced to a linear, first-order initial value problem in Equation 5.13 below. Using this solution (which is assumed to exist in the statement of Theorem 5.1.3), we define

$$g(t, T) = 1 - \int_t^T f(s, T)\alpha_2(s) ds. \quad (5.11)$$

The term structure equation, based on the risk-neutral drift of (C) and the postulated bond pricing function, simplifies to

$$g_t(t, T) - f_t(t, T)r - f(t, T)\left(r^2 + r\alpha_1(t) + \alpha_2(t)\right) = r(g(t, T) - f(t, T)r),$$

or, cancelling the r^2 terms and rearranging, to

$$g_t(t, T) - f(t, T)\alpha_2(t) - r\left(f_t(t, T) + f(t, T)\alpha_1(t) + g(t, T)\right) = 0. \quad (5.12)$$

The first two terms are equal to zero, as a straightforward consequence of the definition of $g(t, T)$ in Equation 5.11. The term in the large brackets is also equal to zero — this can be seen by rewriting Equation 5.10 as

$$f_{tt}(t, T) + \frac{\partial}{\partial t}\left(\alpha_1(t)f(t, T)\right) + f(t, T)\alpha_2(t) = 0,$$

and then integrating (with respect to t) both sides from t to T :

$$f_t(T, T) - f_t(t, T) + \alpha_1(T)f(T, T) - \alpha_1(t)f(t, T) + \int_t^T f(s, T)\alpha_2(s) ds = 0.$$

Applying the terminal conditions, and multiplying by -1 , we have

$$1 + f_t(t, T) + \alpha_1(t)f(t, T) - \int_t^T f(s, T)\alpha_2(s) ds = 0,$$

and, substituting in Equation 5.11, we get

$$f_t(t, T) + \alpha_1(t)f(t, T) + g(t, T) = 0,$$

which shows that Equation 5.12 is satisfied, as required. Finally, the assumed terminal condition $f(T, T) = 0$ and the terminal value $g(T, T) = 1$ (implied by Equation 5.11), give the correct terminal condition $P(T, T, r) = 1$, for all r and $T > 0$. □

In summary, according to Theorem 5.1.3, (1,1) models necessarily exhibit bond prices that are

affine in their short rate, and these models are attained if, and only if, the short rate satisfies a certain class of SDEs.

With regard to the result that (A) implies (B), [Collin-Dufresne & Goldstein \[2002\]](#) arrive at Equation 5.7, and prematurely conclude that such an equation cannot hold for all maturity times. They thus neglect the possibility of affine bond prices, which is an extremely simple way of satisfying Equation 5.7. Another way our result can be expressed is that we have delineated the exact category of counterexamples to their proposition as it is stated. Strictly speaking, it remains to show that this category is non-empty — we do this in Section 5.2.1, by showing that a certain example of these (1,1) models exists. While an affine bond pricing function — not to be confused with the exponential-affine functions involved in affine term structure models — is not standard, this result shows that it is necessary if one is to attain a (1,1) model.

The result that (C) implies (B) tells one how to construct (1,1) models — the short rate must be modelled with the risk-neutral drift function given in (C). Section 5.2 takes up this task. This result is a generalisation of a result implicit in [Filipović et al. \[2017\]](#). Interestingly, their *linear-rational* modelling framework captures a large class of (1,1) models; namely, those that are time homogeneous and have a certain specification of the volatility process $\{u_t\}$ (in particular, the drift coefficient $\mu^{\mathbb{Q},u}(\cdot, \cdot, \cdot)$ must be of a certain form). The calculations and details are given in Appendix D.1. For these cases, the result that (C) implies (B) — the form of the bond pricing function — is implicit in their work. In addition to generalising their results, we also contribute by showing, in Appendix D.1, how linear-rational models function at the level of the short rate under the risk-neutral measure (as they are specified in terms of the pricing kernel under the real-world measure). One benefit of this is that it allows linear-rational models to be compared to the short-rate models that pervade the literature. Other benefits, and the general comparison between our results and those of [Filipović et al. \[2017\]](#), are discussed further in Section 5.2.2.

The result that (B) implies (C) — which ensures that a risk-neutral, quadratic drift for the short rate is the unique method of attaining (1,1) models — was proved by [Gabaix \[2009\]](#) in a time-homogeneous setting.⁹ Our extension to time-inhomogeneous models is noteworthy, as the analogue extension in affine term structure models does not hold — the well-known if-and-only-if equivalence between exponential-affine bond prices and affine short-rate dynamics only holds in a time-homogeneous setting; exponential-affine bond prices do not necessitate affine dynamics in general [[Björk, 2004](#), Ch.24].

The assumption of a bivariate model, with one state variable driving volatility, is not necessary for the (B)-if-and-only-if-(C) aspect of the theorem. In Section 5.3, we briefly consider some extensions and changes of framework in this vein.

The affine-coefficient functions $f(\cdot, \cdot)$ and $g(\cdot, \cdot)$, which necessarily govern bond prices in a (1,1) model, can be further characterised as follows.

Corollary 5.1.5. Consider a (1,1) model (which must satisfy conditions (B) and (C) in Theorem 5.1.3). Then for each $T > 0$, $f(\cdot, T)$ is equal to the first component of the two-dimensional vector-valued function $y^T : [0, T] \rightarrow \mathbb{R}^2$ that satisfies the following first-order, linear ODE system:

$$\frac{\partial y^T(t)}{\partial t} = \begin{bmatrix} 0 & 1 \\ -\alpha_2(t) - \frac{\partial \alpha_1(t)}{\partial t} & -\alpha_1(t) \end{bmatrix} y^T(t) \quad (5.13)$$

⁹This is not done directly: it is shown that a linearity-generating process implies a quadratic drift, and that if the bond price is an affine function then it comes from a linearity-generating process. [Cheridito & Gabaix \[2008\]](#) and [Carr et al. \[2011\]](#) consider a short-rate processes satisfying (C), but do not show that this implies an affine bond pricing function. Note also that while the time-homogeneous setting in [Gabaix \[2009\]](#) is less general, driving martingales more general than Brownian motion are considered.

for all $t \in [0, T]$, with terminal condition $y^T(T) = [0 \ -1]^\top$.

Furthermore, if $\alpha_1(\cdot)$ and $\alpha_2(\cdot)$ have constant values $\alpha_1, \alpha_2 \in \mathbb{R}$ for all $t \in [0, S]$, and if the quadratic $\lambda^2 + \alpha_1\lambda + \alpha_2$ admits two distinct, real roots λ_1 and λ_2 (with $\lambda_1 < \lambda_2$), then

$$f(t, T) = \frac{1}{\lambda_2 - \lambda_1} (e^{-\lambda_1(T-t)} - e^{-\lambda_2(T-t)}) \quad (5.14)$$

and

$$g(t, T) = \frac{1}{\lambda_2 - \lambda_1} (\lambda_2 e^{-\lambda_1(T-t)} - \lambda_1 e^{-\lambda_2(T-t)}) \quad (5.15)$$

for all $t \in [0, T]$.

Proof. Given that condition (A) of Theorem 5.1.3 applies (i.e., that we are considering a (1, 1) model), the bond pricing function must take on the affine form in (B) with, for each $T > 0$, $f(\cdot, T)$ satisfying the second-order ODE in Equation 5.10. Define

$$h(t, T) = f_t(t, T) \quad \text{and} \quad y^T(t) = \begin{bmatrix} f(t, T) \\ h(t, T) \end{bmatrix}$$

for $t \in [0, T]$. This definition gives the first ODE of the system in Equation 5.13, while the second comes from rewriting Equation 5.10 as

$$h_t(t, T) = -h(t, T)\alpha_1(t) - f(t, T) \left(\frac{\partial \alpha_1(t)}{\partial t} + \alpha_2(t) \right).$$

The terminal conditions come directly from the old ones ($f(T, T) = 0$ and $f_t(T, T) = -1$) and the definition of $y^T(\cdot)$ above.

In the case of constant α_1 and α_2 , the ODE system reduces to

$$\frac{\partial y^T(t)}{\partial t} = \begin{bmatrix} 0 & 1 \\ -\alpha_2 & -\alpha_1 \end{bmatrix} y^T(t). \quad (5.16)$$

Provided the characteristic polynomial of the above coefficient matrix ($\lambda^2 + \alpha_1\lambda + \alpha_2$) has real roots λ_1 and λ_2 , ordered so that $\lambda_1 < \lambda_2$, it is easily verified that $f(\cdot, T)$ from Equation 5.14 gives rise to $y^T(\cdot)$ that satisfies Equation 5.16, as well as the two terminal conditions (see Teschl [2012, Ch.3] for details on linear ODE solutions). Straightforward integration, as per Equation 5.11, yields

$$g(t, T) = 1 - \frac{\alpha_2}{\lambda_2 - \lambda_1} \left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2} - \frac{e^{-\lambda_1(T-t)}}{\lambda_1} + \frac{e^{-\lambda_2(T-t)}}{\lambda_2} \right),$$

which can be simplified — based on the roots of the characteristic polynomial — to agree with Equation 5.15. □

We will hereafter focus on (1,1) models conforming to the latter part of Corollary 5.1.5; that is, to models where the quadratic \mathbb{Q} -drift of the short rate has constant coefficients and admits two real roots:

$$\mu^{\mathbb{Q},r}(t, r, u) = \mu^{\mathbb{Q},r}(r) = r^2 + \alpha_1 r + \alpha_2 = (r - \lambda_1)(r - \lambda_2). \quad (5.17)$$

Furthermore, we will parameterise our models in terms of these roots λ_1 and λ_2 (which will turn out to have natural interpretations), making the solutions in Equations 5.14 and 5.15 directly

applicable.

A straightforward condition for avoiding mathematical problems in the short-rate process can now be seen — the quadratic in Equation 5.17 will violate the linear growth condition, and result in an explosion of the short rate, if the process is not bounded above. In order to not contradict the assumption that the state variables are well-defined (Assumption 2.1.1), the short rate will have to be bounded above by λ_2 , so that the increasingly positive drift on the right-hand side of the parabola in Equation 5.17 is avoided.¹⁰

Furthermore, Section 5.2.3 will demonstrate that the atypical affine bond pricing function is in fact suitably behaved and well controlled by the values of λ_1 and λ_2 . Note also that the condition $r_t < \lambda_2$ ensures, in the context of the affine bond pricing function implied by Theorem 5.1.3 and Corollary 5.1.5, that ZCB prices are positive.

5.2 The LADQ(1,1) Model

The goal of this section is to propose a parametric (1,1) model, based on the (specialised) framework of Section 2.1 and on Theorem 5.1.3. We propose the model initially under the risk-neutral measure, and examine the specification in a number of subsections below.

5.2.1 Risk-neutral Specification

We propose the following parametric special case of the dynamics in Assumption 5.0.3:

$$dr_t = (r_t - \lambda_1)(r_t - \lambda_2) dt + r_t(\bar{r} - r_t)u_t dW_t^{(1)} \quad \text{and} \quad (5.18)$$

$$du_t = \kappa(\theta - u_t) dt + \sigma\sqrt{u_t}(\rho dW_t^{(1)} + \sqrt{1 - \rho^2} dW_t^{(2)}), \quad (5.19)$$

where $\{W_t^{(1)}\}$ and $\{W_t^{(2)}\}$ are the two scalar components of $\{W_t\}$, and where $\lambda_1, \lambda_2, \bar{r}, \kappa, \theta, \sigma$ and ρ are constants. This specification conforms to condition (C) in Theorem 5.1.3; the resultant model therefore exhibits a (1,1) structure and an affine bond price, and therefore log-affine yields. Because of this, and because of the quadratic forms of the short rate's drift *and* volatility, we term this the *Log-Affine Double Quadratic (1,1)* — or LADQ(1,1) — model. We now analyse this risk-neutral specification; a real-world measure specification is given below in Section 5.2.5.

The following restrictions are enforced on the short rate process:

$$0 < \lambda_1 < \bar{r} < \lambda_2 \quad \text{and} \quad (5.20)$$

$$0 < r_0 < \bar{r}. \quad (5.21)$$

The short-rate volatility function $\sigma^r(t, r, u) = \sigma^r(r, u) = r(\bar{r} - r)u$ bounds the short rate, ensuring in particular that $r_t < \lambda_2$ for all $t \in [0, S]$. This is simply because if ever we have $r_t = \bar{r}$, the process has zero volatility ($\sigma^r(\bar{r}, u) = 0$) but negative drift (as we have forced, in Equation 5.20, \bar{r} to be in the negative region of the drift's parabola) — \bar{r} is thus an upper bound for the short rate.¹¹ The left-hand panel of Figure 5.1 illustrates this. The volatility also keeps the short rate bounded from below by zero — if for some t we have $r_t = 0$, the short rate becomes locally deterministic

¹⁰Given that the short rate is a state variable, this will prove to be relatively straightforward. Cheridito & Gabaix [2008], as their primary undertaking, investigate how this can be achieved when the short rate is the sum of several processes.

¹¹If we let $\bar{r} = \lambda_2$ it would not be clear whether this would bound the process (as the drift at $r_t = \lambda_2$ is zero). If $\{u_t\}$ was a constant process, the Feller [1951] condition — which pertains to one-dimensional diffusions — would be informative, but the introduction of $\{u_t\}$ leaves the matter open. It is similarly open as to whether the upper bound \bar{r} is attainable.

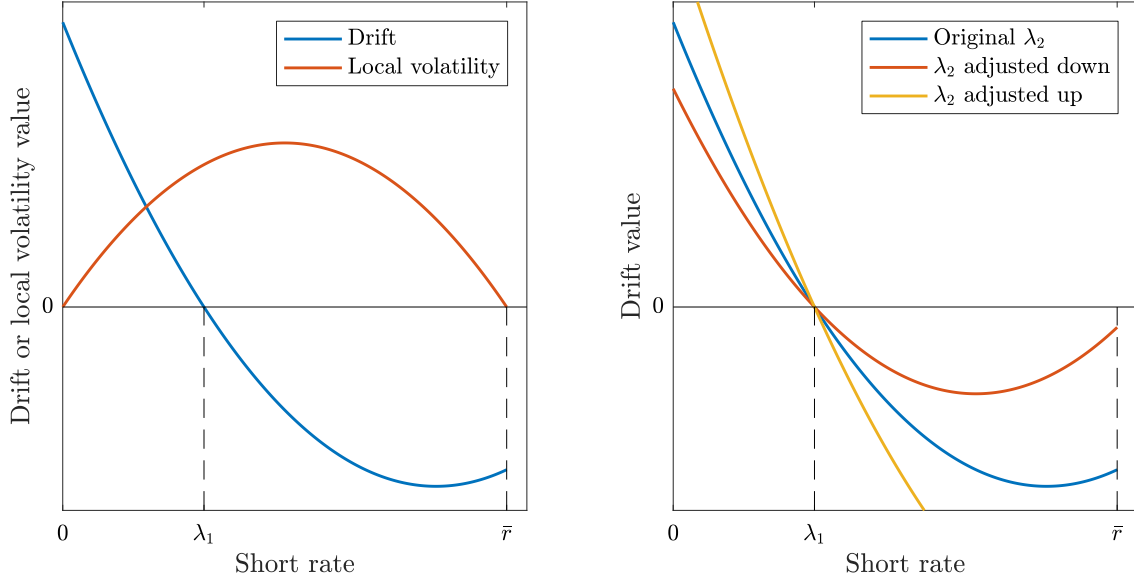


Figure 5.1: In both panels, the (risk-neutral) short-rate drift is plotted over the domain of the short rate. Various values of λ_2 are considered on the right-hand side. The left-hand panel also plots short-rate local volatility, which is equal to the whole volatility coefficient if $u_t = 1$.

(as $\sigma^r(0, u) = 0$), but has a positive drift (of magnitude $\lambda_1 \lambda_2$). Note that these arguments rely on the continuity of $\{r_t\}$ — in order for the short rate to leave the region $[0, \bar{r}]$, it must first equal 0 or \bar{r} .

Given this bounding, one can easily verify that the short rate's dynamics comply with the usual linear growth and Lipschitz conditions — indeed, one can see from Figure 5.1 that, ignoring the influence of the $\{u_t\}$, the short rate's coefficient functions are bounded and differentiable. Once the Feller condition is enforced on the dynamics of $\{u_t\}$ (which is done in Equation 5.22 below), the coupled dynamics satisfy the linear growth and Lipschitz conditions, and thus pose no potential contradiction with the assumption that the state process is well-defined. Provided that one relates \mathbb{Q} and \mathbb{P} with a Radon-Nikodým derivative process that is a martingale (not merely a local martingale), it follows that Assumption 2.1.1 will also be respected, and also that \mathbb{Q} and \mathbb{P} are equivalent. To fully satisfy Assumption 2.1.4, it remains to verify that cash-account-denominated ZCB-price processes are martingales. Itô's lemma (and the satisfaction of the term structure equation) implies that

$$d\left(\frac{P_{tT}}{B_t}\right) = \frac{-f(t, T)r_t(\bar{r} - r_t)u_t}{B_t} dW_t^{(1)}.$$

Although Andersen [2011] shows that local martingales with quadratic volatility functions are often not true martingales, the boundedness of the ZCB process, inherited from the boundedness of the short rate, ensures martingality in this case.

In addition to bounding the process, short-rate volatility function gives the short rate a *local-volatility* structure, to accompany the unspanned volatility component $\{u_t\}$. If we ignore the unspanned volatility by setting u_t to a constant, the short rate's volatility is a negative quadratic in r — see the left-hand panel of Figure 5.1. A stylised fact of empirical yield curve variation is *positive* dependence of interest-rate volatility on rate levels [Piazzesi, 2010, §7.7]. An important model validation step is therefore to verify that the permissible range for \bar{r} in Equation 5.20 is wide enough — to ensure this positive level dependence, \bar{r} would need to be set large enough that

the short rate tends to remain in the region $[0, \bar{r}]$, where the local volatility is increasing.¹²

The short rate process drift function $\mu^{\mathbb{Q},r}(t, r, u) = \mu^{\mathbb{Q},r}(r) = (r - \lambda_1)(r - \lambda_2)$, over its domain $r \in [0, \bar{r}]$, is plotted several times in Figure 5.1. Given the restriction in Equation 5.20, and given the short rate's domain, the drift function makes the short rate *mean reverting* to λ_1 : if $r < \lambda_1$, then $\mu^{\mathbb{Q},r}(r) > 0$; if $r > \lambda_1$, then $\mu^{\mathbb{Q},r}(r) < 0$. The *rate of mean reversion* is controlled by λ_2 — see the right-hand panel of Figure 5.1, where the drift function is replotted for differing values of λ_2 . For a larger value of λ_2 , the drift toward the mean, whether positive or negative, is larger in absolute value. Note that the typical short-rate rate-of-mean-reversion parameter — based on a linear mean-reverting drift term, like κ for the process $\{u_t\}$ above — is not directly comparable to λ_2 , but is best compared to $\lambda_2 - \lambda_1$ (the negative of the slope of our drift function at its intercept).¹³

Concerning the volatility process $\{u_t\}$ defined by Equation 5.19, in addition to enforcing $\rho \in (-1, 1)$, we require that

$$\min(\kappa, \theta, \sigma, u_0) > 0 \quad \text{and} \quad 2\kappa\theta > \sigma^2, \quad (5.22)$$

which keeps the process strictly positive, and reverting to a positive long-term mean (as well as ensuring that σ is identifiable). The process simply scales the volatility of the short rate. Its general level is reflected (under the risk-neutral measure) by θ ; a small value for θ , for instance, implies a local volatility that is less level dependent and smaller in general magnitude. While the local volatility structure is clearly controlled by \bar{r} , the value of θ thus determines how the local volatility manifests.

The degree to which the unspanned volatility component is emphasised over local volatility is dictated by the values of σ and κ . These, of course, control the size of shocks to $\{u_t\}$ and the persistence of these shocks (under \mathbb{Q}), respectively. Larger and more persistent shocks will cause, other things equal, the unspanned volatility to dominate the local volatility.

The correlation between the two independent Brownian motions is written explicitly in Equation 5.19. While level dependence of the local volatility is controlled by θ and \bar{r} , ρ introduces and controls a level dependence of the unspanned volatility component: changes in the short rate will be correlated with changes in unspanned volatility. The LADQ(1,1) model thus has level dependent volatility in two senses.

5.2.2 Comparison to Other Models

Amongst the short-rate modelling literature, the LADQ(1,1) model appears most similar to the *3/2 model* of Ahn & Gao [1999]. The drift functions of both models are quadratic, and both employ a non-standard volatility function: our volatility is quadratic; theirs is proportional to their short rate raised to the power of one and a half.¹⁴ However, this similarity is superficial. Their quadratic drift is parameterised quite differently, and is motivated by empirical concerns; ours is constructed in light of Theorem 5.1.3 (and, as discussed above, achieves a typical mean-reversion mechanism nevertheless). The LADQ(1,1) volatility function (at least its local-volatility aspect) more closely resembles that in the Cox et al. [1985] model — volatility increases with the short rate, from zero, in a concave manner — than that in the 3/2 model. Our parameter θ , which describes the scaling of our local volatility term, is like the volatility parameter in the Cox et al. [1985] model; \bar{r} affords

¹²See Footnote 13.

¹³The estimates for this mean reversion vary significantly in the literature. The lowest we can find is 0.063 [Duffee & Stanton, 2012], based on which the range for \bar{r} becomes $(\lambda_1, \lambda_1 + 0.063)$. This is likely sufficiently wide to ensure the positive level dependence discussed above.

¹⁴Short-rate models' volatility functions are usually proportional to the short rate raised to the power of zero (as in the Vasiček [1977] model), a half (e.g., the Cox et al. [1985] model) or one (in, e.g., the Dothan [1978] model). All of the short-rate models in Brigo & Mercurio [2007] feature these standard exponent values.

additional flexibility over the shape of the local volatility function (large \bar{r} values make our local volatility, in the relevant region, almost linear, like in the [Dothan \[1978\]](#) model).

The direct scaling of the local volatility with the unspanned volatility process $\{u_t\}$ differs from most of the literature. A square root is often taken — see, for instance, the [Fong & Vasicek \[1991\]](#) model shown in Section 1.1 — to ensure the model is affine; we have no such need.

Quadratic volatility structures, like that in Equation 5.18 — have been employed by [Rady \[1997\]](#) and [Ingersoll \[1997\]](#), to model asset price ratios and exchange rates, respectively, in a bounded way. It is also employed, although less explicitly, by [Carr et al. \[2011\]](#): they do not work with the short rate directly, but show that the short-rate volatility function implicit in their model is a negative quadratic, giving them the ability to bound their short rate.

A crucial comparison is to the (1,1) models that pre-exist the LADQ(1,1) model. Appendix D.1 delineates the class of (1,1) models that arise in the linear-rational framework of [Filipović et al. \[2017\]](#).¹⁵ Linear-rational square-root models are introduced in Equation 4.1 in Chapter 4; the case with one spanned and one unspanned state variable gives the LRSQ(1,1) model. Appendix D.2 shows that the short rate in the LRSQ(1,1) model satisfies

$$dr_t = (r_t - \alpha)(r_t - \alpha - k) dt + \frac{(r_t - \alpha - k)^2}{k(1 + \theta_Z)} \sqrt{\frac{\sigma_1^2(\alpha - k\theta_Z - r_t)}{r_t - \alpha - k} + (\sigma_2^2 - \sigma_1^2)u_t} dW_t^{(1)} \quad (5.23)$$

and

$$du_t = \left(k(\theta^{(2)} - u_t) - \frac{u_t \sigma_2^2 (r_t - \alpha - k)}{k(1 + \theta_Z)} \right) dt + \sigma_2 \sqrt{u_t} \left(\rho_t dW_t^{(1)} + \sqrt{1 - \rho_t^2} dW_t^{(2)} \right), \quad (5.24)$$

where $\alpha, k, \theta_Z, \sigma_1, \sigma_2$ and $\theta^{(2)}$ are constants, and where

$$\rho_t = \frac{\sigma_2 \sqrt{u_t}}{\sqrt{\frac{\sigma_1^2(\alpha - k\theta_Z - r_t)}{r_t - \alpha - k} + u_t(\sigma_2^2 - \sigma_1^2)}}.$$

The LRSQ(1,1) model has many strengths.¹⁶ It allows an efficient Fourier-based derivative pricing method, whereas the LADQ(1,1) needs to rely on relatively slow finite-difference methods (see Section 5.2.4 below). It also avoids the explosion problem in a natural, endogenous way (the short rate is given by a suitably bounded function of the original state variables — see Appendix D.2), while the LADQ(1,1) requires an ad hoc upper bound on the short rate. The two bounding mechanisms turn out to be similar — note that the LRSQ(1,1) volatility goes to zero at $r_t = \alpha + k$ (which is clearly the analogue to λ_2).

Working in a short rate setting has benefits, in addition to the ability to compare the LRSQ(1,1), LADQ(1,1) and other short-rate models.¹⁷ Firstly, Theorem 5.1.3 shows that (1,1) are naturally characterised in terms of their short rate — one can now see precisely how the LRSQ(1,1) and LADQ(1,1) relate to the whole class of (1,1) models. Secondly, the economic concreteness of the short rate is helpful. For example, we are easily able to ensure our volatility specification conforms to the positive level dependence that is typically observed. Also, because the short rate is a spanned state variable, one can look at Equation 5.18 or 5.23 and easily discriminate the local and

¹⁵We do not address of (1,1) model of [Carr et al. \[2011\]](#) firstly because [Filipović et al. \[2017\]](#) show that it can be incorporated into the linear-rational framework, and secondly because it is time-inhomogeneous, making it difficult to compare to the LADQ(1,1) model and unsuitable for many applications.

¹⁶Note that we are not discussing general linear-rational models, which have many relative advantages, such as the ability to easily incorporate USV and to tractably apply Fourier methods to swaption pricing.

¹⁷We also note that we appear to be one of only a few short-rate models with USV — the only others known to the author are the affine term structure models outlined by [Collin-Dufresne & Goldstein \[2002\]](#) and generalised by [Joslin \[2017\]](#). Note also that affine models cannot simultaneously exhibit USV and non-negative interest rates [[Filipović et al., 2017](#)].

unspanned volatility components. Thirdly, the market price of risk — the link between the risk-neutral and real-world probability measures — is much easier to control in a short rate setting. In Section 5.2.5, we simply specify a suitable market-price-of-risk process; in the linear-rational and related frameworks, the market price of risk is endogenous to the specification of the pricing kernel process. Filipović et al. [2017] need to introduce an auxiliary probability measure to augment the endogenous specification.

The LADQ(1,1) model has some advantages over the LRSQ(1,1) model. Firstly, it ensures non-negative interest rates (or any other lower bound — see Section 5.3.1). Secondly, it allows a simple and effective parametric control of the yield curve, described below in Section 5.2.3. The LRSQ(1,1) model can only enjoy one of these two benefits. Filipović et al. [2017] recommend setting $\alpha = k\theta_Z$, which ensures — see Appendix D.2 — non-negative rates. This, however, sacrifices using α (the analogue of λ_1) to freely control the yield curve. Thirdly, in Section 5.2.4 we describe how the LADQ(1,1) model affords a straightforward control over the height, slope and curvature of implied-volatility skews.

5.2.3 LADQ(1,1) Term Structure

The parametric form of the LADQ(1,1) yield curve is implied by Theorem 5.1.3 and Corollary 5.1.5: the time- t continuously compounded yield on a T -dated ZCB is given by

$$y_{tT} = -\frac{\log P_{tT}}{T-t} = -\frac{1}{T-t} \log (g(t, T) - r_t f(t, T)),$$

where $f(t, T)$ and $g(t, T)$ are specified by Equations 5.14 and 5.15.

It is shown in Section 5.2.1 that the short rate reverts to λ_1 under the risk-neutral measure. One would therefore expect the yield of a long-dated bond to be λ_1 ; indeed, it is easy to verify that

$$\lim_{T \rightarrow \infty} y_{tT} = \lambda_1.$$

The level of the long end of the yield curve is thus given by λ_1 (see the left-hand panel of Figure 5.2), while the short end is given by the prevailing short rate. Because, as described above, λ_2 controls the rate of (risk-neutral) mean reversion, it should also dictate how fast short-term yields converge to the long end; that is, λ_2 should control the yield curve's curvature. The right-hand panel of Figure 5.2 confirms this.

With regard to parametric control of the yield curve then, the LADQ(1,1) model turns out to be similar to the classical Vasiček [1977] and Cox et al. [1985] models — despite the short rate's non-standard dynamics, the yield curve's qualitative features (short-term yields, long-term yields and rate of convergence from the former to the latter) are straightforwardly controlled. In Section 5.2.6, we consider adding a time dependency to the model, so that $\{y_{0T}\}_{T \geq 0}$ can match an exogenously specified yield curve.

5.2.4 LADQ(1,1) Option Pricing

Because of the unspanned nature of the volatility process, the ability to price derivative instruments is crucial — realisations of $\{u_t\}$ cannot be inferred from yield curve observations, and so must be linked to prices of volatility-sensitive derivatives. The non-standard dynamics in Equation 5.18 preclude the use of Fourier-based techniques, which require the characteristic function. We therefore opt for a finite-difference method, which performs well in low-dimensional environments.

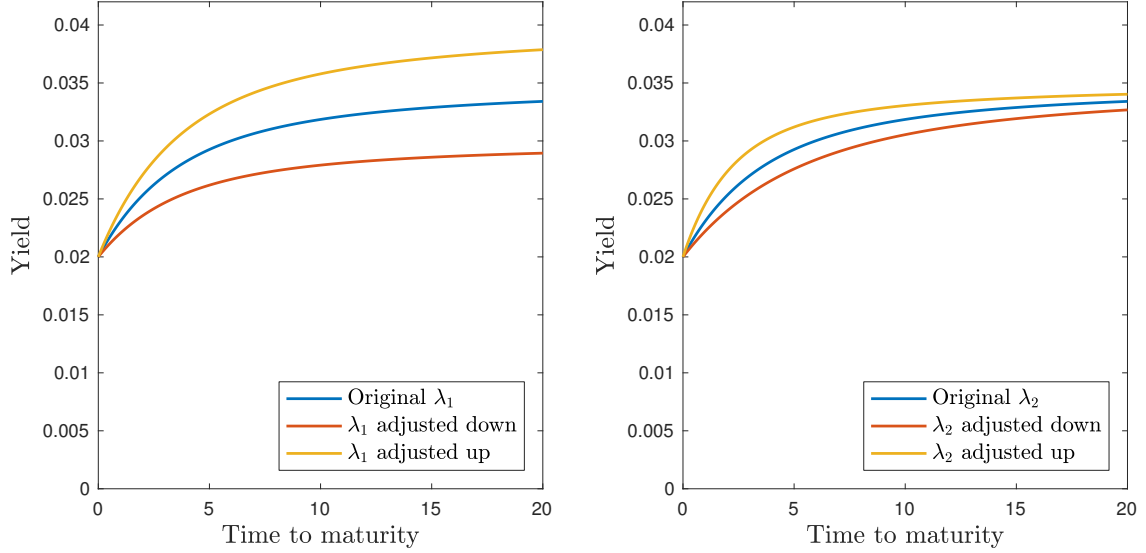


Figure 5.2: The LADQ(1,1) term structure is plotted in both panels, supposing the short rate is 2%. The left- and right-hand panels vary λ_1 and λ_2 , respectively.

Suppose a derivative pays out $\Phi(r_T)$ at time T . The no-arbitrage time- t price of such a claim must, due to the Markov nature of the state variables, be given by some function $p_t = p(t, r_t, u_t)$. As is well known, the claim process $\{p_t\}$ must offer an expected rate of return, under \mathbb{Q} , equal to the short rate. The expected rate of return can be computed with Itô's Lemma and the LADQ(1,1) dynamics in Equations 5.18 and 5.19. The resulting PDE is paired with the terminal condition $p_T = \Phi(r_T)$. Having discretised the time and state-variable domains, and truncated the volatility domain with some upper bound \bar{u} , we employ an alternating-direction implicit (ADI) scheme (outlined originally in Peaceman & Rachford [1955] and subsequently in, among others, Crépey [2013]).

The LADQ(1,1) closed-form bond price allows typical derivatives such as bond options, caps or swaptions (or indeed any European option depending on the yield curve) to be explicitly represented as T -dated claims that depend on the short rate.¹⁸ Consider, for example, an at-the-money put option written on the two-year ZCB, with expiry in one year's time, which pays out $(K^{\text{ATM}} - P_{12})^+ = (\frac{P_{02}}{P_{01}} - P_{12})^+ = (\frac{P_{02}}{P_{01}} - g(1, 2) + f(1, 2)r_1)^+$ at $T = 1$. The left-hand panel of Figure 5.3 plots the finite-difference pricing surface, while the right-hand panel fixes an initial volatility value. We verify our finite-difference method by comparing its estimates, in the right-hand panel, to Monte-Carlo estimates; in particular, we plot the 99% Monte-Carlo confidence interval for the put option price.¹⁹

A virtue of the LADQ(1,1) is that certain parameters are straightforwardly related to qualitative

¹⁸Swaptions (and coupon-bond options) are option problematic to price because of the pay out's dependence on various points on the yield curve. In one-factor models, the decomposition technique of Jamshidian [1989] can solve this difficulty by expressing the swaption as a portfolio of simpler, ZCB options. This is applicable in a (1,1), or indeed a (1, n), model. However, the affine bond form makes this unnecessary — for instance, the pay out of a standard payer swaption is proportional to

$$\left(\sum_{i=1}^n \delta P_{TT_i} + P_{TT_n} - 1 \right)^+,$$

which, in the LADQ(1,1) or indeed any (1,1) model, can be written as an affine function of the short rate. There is thus no need to implement the Jamshidian [1989] decomposition.

¹⁹Specifically, we use an Euler-Maruyama discretisation of the dynamics in Equations 5.18 and 5.19, with 200 times steps for the one year. We implement 5000 such paths for each initial short rate value, and estimate the discount factor with a trapezoidal rule. The interval is defined as the mean of the realised discounted pay outs, shifted up and down by three times the estimated standard deviation of this mean.

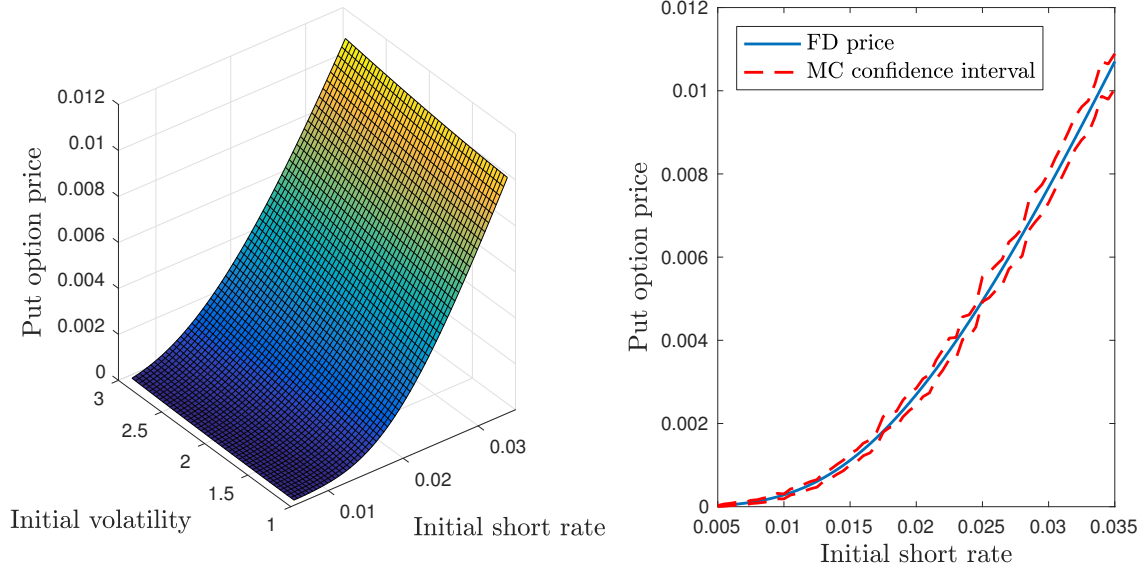


Figure 5.3: Finite-difference (FD) prices of one-year, at-the-money put options, written on the two-year ZCB, are plotted in both panels. The left-hand panel illustrates the whole pricing-function surface. The right-hand panel fixes $u_0 = 2$, and plots prices for different initial values for r_0 (without changing the initially determined strike price), and also plots of the boundaries of a Monte-Carlo (MC) estimated 99% confidence interval for the same put option price.

features of the volatility skew: like in the SABR model, θ , ρ and σ control the skew's height, slope and curvature, respectively [Hagan, Kumar, Lesniewski & Woodward, 2002] — see Figure 5.4. With regard to the term structure of implied volatilities, we have more relevant parametric freedom than in the SABR model. Like in the Heston [1993] model, θ and u_0 can be used to control the long and short volatilities, respectively, with κ dictating the speed of convergence between them.

Finite-difference methods are slower than semi-analytical Fourier-based pricing methods (available in the LRSQ(1,1) model). However, two commonly applicable efficiencies mitigate this significantly. Firstly, when pricing a long-dated option, the intermediate solutions provide the prices of similar, but shorter-dated options. In the important case of a cap struck at some particular rate, one only needs to price the longest of the constituent caplets — the remaining caplet prices can be found in the intermediate solutions and require no further computation. Secondly, because the finite-difference scheme outputs the whole pricing function, one does not need to recompute option prices for different values of the state variables. This can be extremely useful when estimating the model from panel data: when attempting a particular parameter set, certain options need only be priced once, as the same pricing surface is applicable as the state variable values change over the time series.²⁰ An additional benefit of our method is the ability to easily incorporate early exercise features.

5.2.5 LADQ(1,1) Time-series Dynamics

We now equip the LAQD(1,1) model with real-world measure dynamics. We simply need to specify an adapted, two-dimensional market-price-of-risk process $\{\zeta_t\}$, which relates \mathbb{Q} to the real-world

²⁰This is applicable to options of a fixed strike; the changing strike of at-the-money options would require repricing over the time series.

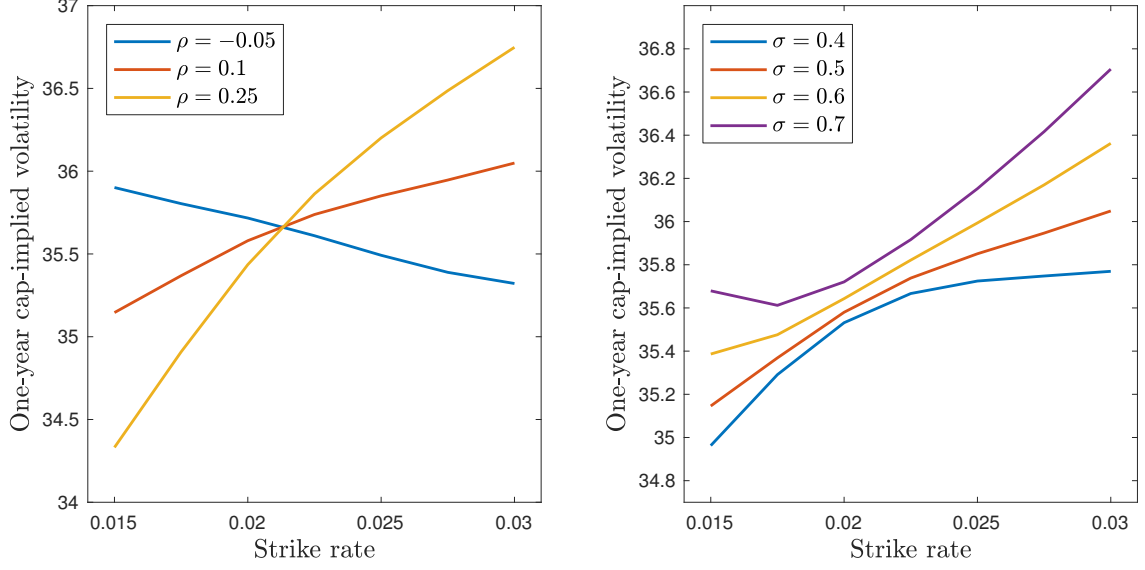


Figure 5.4: Cap-implied volatility skews under the LADQ(1,1) model are plotted. The left- and right-hand panels vary ρ and σ , respectively, resulting in varied slopes and curvatures of the skew.

measure \mathbb{P} (as per Equation 2.3). We do so with

$$\zeta_t^{(1)} = \frac{r_t(\lambda_1 + \lambda_2 - \delta_1 - \delta_2) - \lambda_1\lambda_2 + \delta_1\delta_2}{r_t(\bar{r} - r_t)u_t}$$

and

$$\zeta_t^{(2)} = \frac{\kappa^{\mathbb{P}}\theta^{\mathbb{P}} - \kappa\theta + u_t(\kappa - \kappa^{\mathbb{P}}) - \sigma\sqrt{u_t}\rho\zeta_t^{(1)}}{\sigma\sqrt{u_t}\sqrt{1 - \rho^2}}$$

for all $t \in [0, S]$, for constant parameters δ_1 , δ_2 , $\kappa^{\mathbb{P}}$ and $\theta^{\mathbb{P}}$. This yields the following real-world dynamics:

$$dr_t = (r_t - \delta_1)(r_t - \delta_2) dt + r_t(\bar{r} - r_t)u_t dW_t^{\mathbb{P},(1)} \quad \text{and} \quad (5.25)$$

$$du_t = \kappa^{\mathbb{P}}(\theta^{\mathbb{P}} - u_t) dt + \sigma\sqrt{u_t}(\rho dW_t^{\mathbb{P},(1)} + \sqrt{1 - \rho^2} dW_t^{\mathbb{P},(2)}), \quad (5.26)$$

where $\{W_t^{\mathbb{P},(1)}\}$ and $\{W_t^{\mathbb{P},(2)}\}$ denote the scalar components of the \mathbb{P} -Brownian motion $\{W_t^{\mathbb{P}}\}$. The new parameters δ_1 , δ_2 , $\kappa^{\mathbb{P}}$ and $\theta^{\mathbb{P}}$ play similar roles to their risk-neutral counterparts. While λ_1 was the long-term \mathbb{Q} mean of the short rate (and therefore the level of the long end of the yield curve), δ_1 gives the real-world long-run mean and therefore the typical level of the short rate. Because yield curves tend to slope upward (with risk premia for longer, more volatile bonds), we would very likely have $\delta_1 < \lambda_1$. No such systematic difference between the \mathbb{P} and \mathbb{Q} rates of short-rate mean reversion (δ_2 and λ_2) can be expected. It is unlikely, therefore, that the following requirement will prove restrictive:

$$0 < \delta_1 < \bar{r} < \delta_2.$$

This analogue to Equation 5.20 ensures that the short rate's \mathbb{P} specification is similarly unproblematic.

With Equations 5.25 and 5.26, the LADQ(1,1) model can be applied longitudinally. Based on an Euler-Maruyama discretisation, some short-rate simulation results are shown in Figure 5.5.

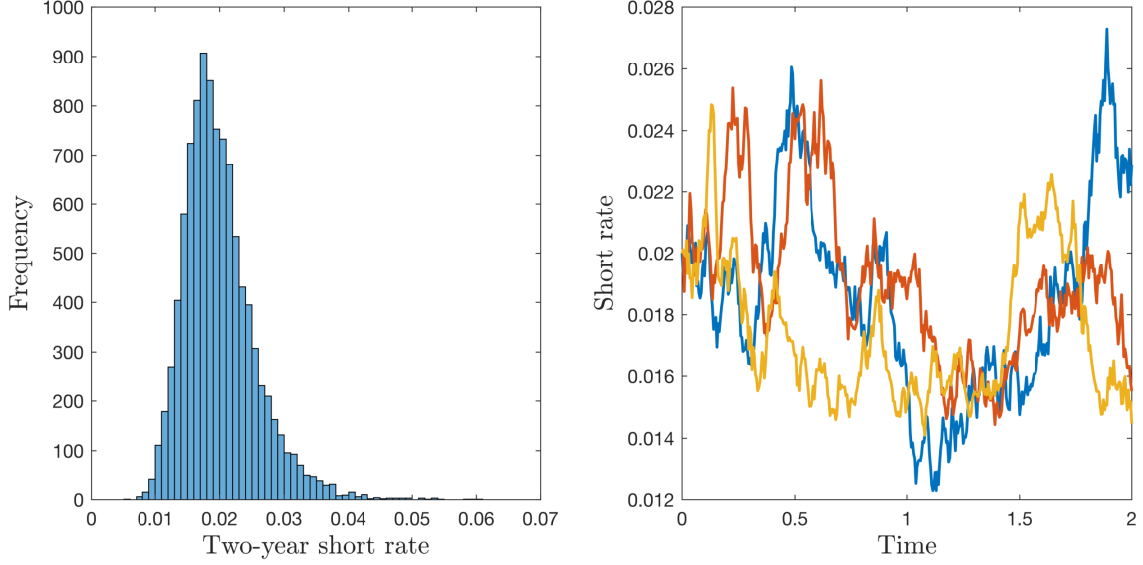


Figure 5.5: The left-hand panel displays a histogram of 10000 realisations of r_2 ; the right-hand panel shows the full path for three of these simulations. The initial short rate r_0 and long-run mean δ_1 were set to 2%.

5.2.6 Yield Curve Calibration

The \mathbb{P} specification is irrelevant for strictly cross-sectional matters such as derivative pricing. Because derivatives are typically hedged with yield-curve instruments, matching the initial term structure of interest rates becomes important. Instead of considering time-dependent λ_1 or λ_2 (allowed by Section 5.1 but not by the latter part of Corollary 5.1.5), we opt for the technique described in Brigo & Mercurio [2001], where the time-homogeneous short rate is *shifted* by the addition of a deterministic function. The original LADQ(1,1) specification, in Equations 5.18 and 5.19, is maintained, but is taken to describe the *unshifted* short rate process, which we denote $\{r_t^*\}$. We then define the actual, shifted short rate as

$$r_t = r_t^* + \varphi(t)$$

for all $t \in [0, S]$. The deterministic function $\varphi(\cdot)$ is chosen so that the model's initial term structure agrees with the market prevailing one; i.e., so that, for all $T > 0$,

$$\begin{aligned} P_{0T}^m &= \mathbb{E}^{\mathbb{Q}}[e^{-\int_0^T r_s^* + \varphi(s) ds}] \\ &= e^{-\int_0^T \varphi(s) ds} \mathbb{E}^{\mathbb{Q}}[e^{-\int_0^T r_s^* ds}] \\ &= e^{-\int_0^T \varphi(s) ds} (g(0, T) - f(0, T)r_0^*), \end{aligned} \quad (5.27)$$

where $\{P_{0T}^m\}$ is the time-0 market-observed term structure of ZCB prices.

Let ZCP_{tTS}^K denote the time- t price of a put option, expiring at T and struck at K , on the ZCB maturing at S (with $t \leq T \leq S$). Brigo & Mercurio [2001] point out that such put options, under the shifted model, can be computed in terms of the unshifted model's prices with

$$ZCP_{tTS}^K = e^{-\int_t^S \varphi(s) ds} \widetilde{ZCP}_{tTS}^{K \exp(\int_t^S \varphi(s) ds)}, \quad (5.28)$$

where \widetilde{ZCP}_{tTS} is the put price under the unshifted model (here, with an adjusted strike). Impor-

tantly, Equation 5.28 is applicable without explicit access to the time-shift function.²¹ Because, following from Equation 5.27, we have

$$e^{-\int_t^S \varphi(s) ds} = \frac{P_{0S}^m \tilde{P}_{0t}}{\tilde{P}_{0S} P_{0t}^m} \quad \text{and} \quad e^{\int_T^S \varphi(s) ds} = \frac{\tilde{P}_{0S} P_{0T}^m}{P_{0S}^m \tilde{P}_{0T}}, \quad (5.29)$$

one can compute the relevant arguments in Equation 5.28, knowing only the market prices P_{0t}^m , P_{0T}^m and P_{0S}^m . Here $\{\tilde{P}_{0T}\}$ denotes the ZCB prices under the unshifted model (i.e., those given by the affine pricing formula).

In order to price caps consistently with a given yield curve then, one should firstly parameterise the time-homogeneous model (by, for instance, setting the r_0 and λ_1 to the prevailing short and long yields, respectively).²² Then one should use the finite-difference method to price the relevant portfolio of ZCB put options as per Equation 5.28 (a cap is equivalent to a portfolio of ZCB puts; this portfolio then needs the adjustments given in Equation 5.28, which in turn make use of Equation 5.29). In order to hedge the cap's short-rate sensitivity with a ZCB, one should note that the ZCB partial derivative with respect to the short rate needs adjustment (the cap's partial derivative can be estimated with a straightforward finite difference) — in the unshifted model, we had $P_r(0, T, r) = -f(0, T)$; from Equation 5.27, one can see we now have

$$P_r(0, T, r) = -f(0, T) e^{-\int_0^T \varphi(s) ds} = -f(0, T) \frac{P_{0S}^m}{\tilde{P}_{0S}}.$$

5.2.7 Empirical Calibration Example

In this subsection, we consider a particular cross-section of the JIBAR data introduced in Section 4.1.2. In addition, caps struck at an additional, in-the-money rate of 5% are included (in-the-money contracts were unsuitable for the concerns of Chapter 4). Section 5.2.4 discussed the LADQ(1,1) model's ability to accommodate various strikes simultaneously — here, we consider various expiries in addition to various strikes (as well as ratify the discussion in Section 5.2.4 with a reduced calibration, across strikes only). In particular, considering the number of parameters available for such a calibration, we include the first three years' caps, so that in total we have nine implied volatilities from the market (caps struck at 5%, 6% and 7%, with final expiries in one, two and three years' time). The LADQ(1,1) risk-neutral parameters (as per Section 5.2.1) are varied so that the model-implied cap-implied volatilities (calculated as per Section 5.2.4) match the market ones as closely as possible (in the sense that the sum squared differences are minimised).

The yield-curve calibration technique of Section 5.2.6, as well as the sequential parameter determination procedure described there, are employed in this process, ensuring that the model is consistent with the market-prevailing yield curve and that the yield-curve parameters conform to the roles described in Section 5.2.3. To avoid an overparameterisation to the term structure of the implied volatilities, we set $u_0 = \theta$ (the pattern of the market quotes over expiries must then be fully accounted for by the natural dynamics of the model, rather than also relying on a fully free value for initial volatility, which should also depend on adjacent cross-sections). We also set $\bar{r} = \lambda_1 - 0.01$, firstly because \bar{r} has little effect on cross-sectional option prices (certainly not a

²¹One could differentiate either side with respect to T (assuming this is possible on the left), and attain $\varphi(\cdot)$ explicitly in terms of the market instantaneous forward rates. This would usually require interpolating between discrete points on the yield curve, and different techniques for doing so result in different functions $\varphi(\cdot)$. Certain applications that we do not consider — such as pricing path-dependent options with Monte Carlo — require that the time-shift function be obtained in this way.

²²One thus pre-calibrates the time-homogeneous model to the extent possible, which seems desirable (one could also set λ_2 to best match the shape of the intermediate yields). In other words, it seems parsimonious to shift the model as little as possible — calibrating to the short and long end ensures that $\varphi(0) = 0$ and $\lim_{T \rightarrow \infty} \varphi(T) = 0$. Brigo & Mercurio [2001] discuss having $\varphi(0) \neq 0$ from a different point of view.

straightforward qualitative one) and is therefore not well revealed in a calibration exercise, and secondly because this gives the widest possible range for the short rate while obeying Equation 5.20 (and is thus conservative — a time-series-based investigation would be needed to motivate a smaller value for \bar{r}).

The resultant fitted implied volatilities are shown in Figure 5.6, and the resultant parameters in the caption. The bottom-right-hand panel presents the results of a reduced calibration, where fewer option prices are taken as calibration conditions; namely, only the two-year caps. For this, κ is set as per the larger calibration exercise, but θ , ρ and σ are allowed to vary.

The calibration is successful — the existence of a global optimum, with parameters in reasonable ranges, was confirmed, and a close fit to the calibration conditions is achieved (especially at the one- and two-year expiries). Although a measure adjustment would be necessary for a detailed interrogation, the volatility parameters indicate the approximate domain in which the volatility process would typically evolve (θ , for example, gives a good sense of the magnitude of the volatility process, while the estimated ρ confirms the level-dependence property discussed in Section 5.2.1). Note also that a negative *risk-neutral* mean reversion is not uncommon to find (see, for instance, Brigo & Mercurio [2007, Ch.3.14]). Furthermore, the LADQ(1,1) fits the sharp increase in the curvature of implied volatility skews as expiry increases — although the calibration errors in the three-year caps are the non-negligible, the model accommodates this rapid increase well.

With regard to the reduced calibration, one would expect there to be sufficient freedom from the three parameters θ , ρ and σ to match the three two-year cap quotes, and indeed — see the bottom-right-hand panel of Figure 5.6 — they do. The contrast between the top-right-hand and bottom-right-hand panels is informative for any future calibration exercise, where the number and structure of the exogenous data must be decided. The comparison between the resulting values for θ , ρ and σ — see the caption of Figure 5.6 — is relatively minor, and does conform to the qualitative roles explained in Section 5.2.4 (for instance, a slightly larger correlation is required to accommodate the relatively large out-the-money quote). A user of the model would need to decide whether parameter changes between the various expiries are worth the increased calibration accuracy — Figure 5.6 illuminates this trade-off.

5.3 Extensions and Conclusion

A more detailed empirical investigation would follow naturally from Section 5.2. The LADQ(1,1) model could be calibrated to a series of volatility skews, and the stability of the resultant parameters examined. Using the real-world measure specification in Section 5.2.5, and also using the numerical efficiencies described in Section 5.2.4, it could be fitted to panel data. While this could lead to a thorough empirical analysis of the LADQ(1,1) model, a particular question of interest would be how the model's total volatility is partitioned between the local and unspanned components.

We conclude the chapter by briefly outlining some extensions to the models, exploiting the fact that bond pricing function in Theorem 5.1.3 and Corollary 5.1.5 depend only on the short rate's risk-neutral drift — the number of state variables can be changed without affecting this important result.

5.3.1 Non-zero Lower Bounds

The LADQ(1,1) model, through the short-rate volatility specification, bounds the short rate below by zero. Because $\lambda_1 > 0$ (Equation 5.20), yields of all maturities are also positive. This can

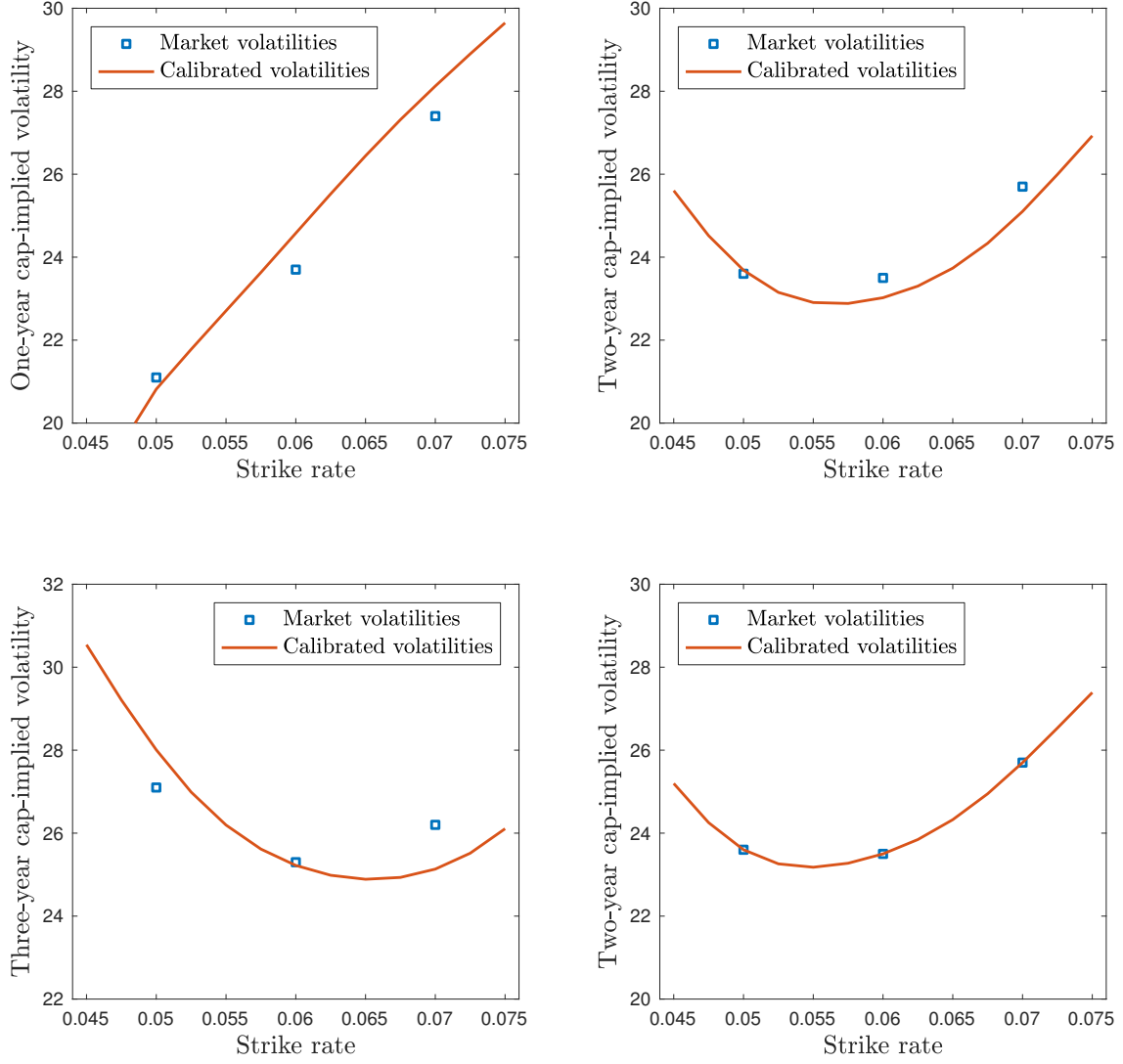


Figure 5.6: The various panels show the model-implied cap-implied volatility as a function of strike price, where the parameters are determined by the described calibration procedure. The market calibration conditions are also shown, as are the results of the reduced calibration procedure, in the bottom-right-hand panel. The parameters resulting from the calibration procedure are: $\lambda_1 = 0.0934$, $\lambda_2 = 0.45$, $\bar{r} = 0.44$, $\kappa = -0.8252$, $\theta = 0.8458$, $\sigma = 0.5479$, $\rho = 0.7372$, $r_0 = 0.055$ and $u_0 = 0.8458$. The parameters that differ in the reduced calibration procedure are: $\theta = 0.8806$, $\sigma = 0.4985$, $\rho = 0.7483$ and $u_0 = 0.8806$.

straightforwardly be generalised:

$$dr_t = (r_t - \lambda_1)(r_t - \lambda_2) dt + (r_t - \underline{r})(\bar{r} - r_t)u_t dW_t^{(1)}$$

for a constant \underline{r} that becomes the short-rate and yield lower bound. One can impose a particular lower bound, or freely estimate a value from data.

One could also allow the lower bound to vary — a third state variable $\{\underline{r}_t\}$ can be introduced to replace \underline{r} above. It could, for instance, satisfy

$$d\underline{r}_t = -\kappa^{\text{LB}} \underline{r}_t dt + \sigma^{\text{LB}} dW_t^{(3)},$$

where κ^{LB} and σ^{LB} are constants and $\{W_t^{(3)}\}$ is a third scalar \mathbb{Q} -Brownian motion. In this specification, $\{\underline{r}_t\}$ will vary around zero (under \mathbb{Q}); other specifications could involve a strictly positive or negative lower bound.

5.3.2 Further Volatility State Variables

Additional volatility state variables can be added to achieve a model with a $(1, n)$ structure. For example, a natural extension of the LADQ(1,1) model is given by

$$dr_t = (r_t - \lambda_1)(r_t - \lambda_2) dt + (r_t - \underline{r})(\bar{r} - r_t)u_t^{(1)}u_t^{(2)} dW_t^{(1)},$$

where $\{u_t^{(1)}\}$ and $\{u_t^{(2)}\}$ are both unspanned volatility processes. If their mean-reversion rates were significantly different, they could be interpreted as modelling short- and long-term volatility shocks separately.

5.3.3 One-factor Models

A class of one-factor term structure models — already outlined by [Gabaix \[2009\]](#) and [Cheridito & Gabaix \[2008\]](#) — follows from Section 5.1:

$$dr_t = (r_t - \lambda_1)(r_t - \lambda_2) dt + \nu(r_t) dW_t,$$

where $\nu(\cdot)$ is a suitable volatility function and $\{W_t\}$ is a scalar \mathbb{Q} -Brownian motion. A specific instance is the LADQ(1,0) model:

$$dr_t = (r_t - \lambda_1)(r_t - \lambda_2) dt + r_t(\bar{r} - r_t) dW_t,$$

which ensures that $r_t \in [0, \bar{r}]$. The three parameters play very similar roles to the three parameters in the classical [Vasiček \[1977\]](#) and [Cox et al. \[1985\]](#) models.

Appendix A

Existence of Suitable State Process Transformations

Following Remark 3.1.15 and the immediately preceding comments, a necessary and sufficient condition for a given (invertible) state process transformation to satisfy Equation 3.9 (the primary requirement of an \mathcal{S} -transformation, as per Definition 3.1.14) is for the final \bar{p} columns of the inverse Jacobian $J_B(\hat{x})$ to lie in $\mathcal{S}^{B(\hat{x})} \hat{X}_*(\lambda \otimes \mathbb{P})$ -almost everywhere (we assume here that there are no partially spanned state variables, so that $\mathcal{X}_S^{(\bar{p})}$ occupies $X_*(\lambda \otimes \mathbb{P})$ -almost all of the state space, i.e., the unspanned state directions \mathcal{S}^x are $X_*(\lambda \otimes \mathbb{P})$ -almost everywhere \bar{p} dimensional).

Remark 3.1.15 notes that if unspanned state directions do not change with the state, a linear \mathcal{S} -transformation can easily be defined. If the unspanned state directions are state dependent, how one should proceed is not obvious. One can, for each \hat{x} , select a particular basis for $\mathcal{S}^{B(\hat{x})}$ as well as additional basis vectors that extend it to a basis for \mathbb{R}^n (as is done in the proof of Lemma 3.1.4), and, using these vectors as columns, write a *candidate* for $J_B(\hat{x})$; however, there does not necessarily exist a transformation $B(\cdot)$ that gives rise to this candidate Jacobian, because the vector-valued function given by each row is not necessarily the gradient of any function (i.e., the *vector field* — the function from \mathbb{R}^n to \mathbb{R}^n — given by each row is not necessarily *conservative*).

One can instead focus on $J_A(\cdot)$ — it is noted in Remark 3.2.12 that if the first $\bar{m} = n - \bar{p}$ rows of $J_A(x)$ (that is, $\nabla A^{(i)}(x)^\top$ for $i = 1, \dots, \bar{m}$) are $X_*(\lambda \otimes \mathbb{P})$ -almost everywhere elements of $(\mathcal{S}^x)^\perp$, then the primary requirement of an \mathcal{S} -transformation will be met.¹ Consider the further requirement that the rows of $J_A(x)$ be mutually orthogonal — then, up to scaling, first \bar{m} rows of $J_A(x)$ would form an orthonormal basis for $(\mathcal{S}^x)^\perp$, and the final \bar{p} rows an orthonormal basis for \mathcal{S}^x .² This would allow a clear and intuitive interpretation: $A^{(i)}(x)$ would have its greatest rate of increase in the direction given by a certain unit-length basis vector that we denote $\delta_i(x)$ (if $i \leq \bar{m}$, $\delta_i(x)$ is a basis vector for $(\mathcal{S}^x)^\perp$, and would be *level* in directions given by the other orthogonal basis vectors $\delta_j(x)$ (that is, it would have a directional derivative of zero in those directions). Thus, the transformed state variable $\{\hat{X}_t^{(i)} = A^{(i)}(X_t)\}$ would track how the movements of the original state process move in the (state-dependent) direction $\delta_i(x)$, and would ignore movements in the other directions $\delta_j(x)$. The final \bar{p} transformed state variables then capture the variation of the state process along unspanned directions (as $\{\delta_i(x)\}_{i=\bar{m}+1}^n$ gives a basis for \mathcal{S}^x), so that the transformed bond pricing function is not sensitive to these state variables.

¹That this is equivalent to the requirement on the final \bar{p} columns of $J_B(\hat{x})$ is simply due to the orthogonality constraints implicit in the relation $J_A(x)J_B(A(x)) = I_n$.

²This would also ensure the *local invertibility* of $A(\cdot)$ (via the inverse function theorem); its global invertibility is discussed later.

This can be summarised by saying that $A(\cdot)$ would give rise to an *orthogonal curvilinear coordinate system* (so that $\hat{X}_t = A(X_t)$ gives the location of the state process X_t in terms of the new coordinates), where the *coordinate curves* of the i th coordinate — the sets of points for which all coordinates (or state variables) except the i th are constant — are everywhere tangent to $\delta_i(x)$. See, for instance, [Lebedev & Cloud \[2003, Ch.4\]](#) for an introduction to curvilinear coordinate systems.

With the aim of characterising such a coordinate system, consider the following first-order ODE system:

$$\frac{dh^{(i)}(r)}{dr} = \delta_i(h^{(i)}(r)), \quad (\text{A.1})$$

where $\delta_i(\cdot)$ is a one of the orthonormal basis vectors (and therefore vector fields) $\{\delta_i(x)\}_{i=1}^n$ described above. A solution of Equation A.1 is a function of real numbers $h^{(i)}(\cdot)$ that returns values in the state space \mathcal{X} . A solution is therefore a *curve* in \mathbb{R}^n , which must furthermore be everywhere tangent to $\delta_i(x)$ (as the curve's rate of change is given, in Equation A.1, by $\delta_i(\cdot)$ evaluated at its location). The solutions of Equation A.1 — for the various possible initial conditions $h^{(i)}(0)$ — are known as the *integral curves* of the vector field $\delta_i(\cdot)$ (see [Lee \[2002, Ch.17\]](#) or [Jost \[2011, Ch.2.2\]](#)). These integral curves give the coordinate curves of the transformed state space (that is, they give the sets of transformed states for which all but the i th transformed state variables are equal). From a mathematical perspective, it is important to note that the length of the vectors in the vector field in Equation A.1 — which we have standardised to one — does not affect the image of integral curves, but only their resulting parameterisation (see, for instance, the discussion of [John \[1982, Ch.1\]](#)).

The sets of coordinate curves are at the heart of the new coordinate system and necessary transformation. All that remains is to suitably index the various positions along each set of integral curves; these index values can then be taken on by $A^{(i)}(\cdot)$, to reflect state movements along an δ_i -integral curve. This must be done in a way that is independent of the particular integral curve occupied — if the state process moves orthogonally to the i th vector field, it moves from one δ_i -integral curve to another. This should not be reflected by $A^{(i)}(\cdot)$, which should be level in such orthogonal directions (this orthogonal movement would be reflected by the indexing of other state variables/coordinates). Assuming the curves are globally well-defined (i.e., that Equation A.1 admits a solutions everywhere in $\mathcal{X}_S^{(\bar{p})}$ — the other $(X_*(\lambda \otimes \mathbb{P})$ -null) states are given separate treatment below), this can be achieved by straightforwardly referring to the parameterisation of *one integral curve*. We leave a detailed treatment of this indexing for future work, but highlight the intuition of the integral-curve-based transformation by reconsidering Example 3.

Example 3 (Continued). A suitable orthonormal basis comes directly from Equation 3.8; we have

$$\begin{bmatrix} \delta_1(x) & \delta_2(x) \end{bmatrix} = \begin{bmatrix} \frac{x^{(1)}}{\sqrt{(x^{(1)})^2 + (x^{(2)})^2}} & \frac{x^{(2)}}{\sqrt{(x^{(1)})^2 + (x^{(2)})^2}} \\ \frac{x^{(2)}}{\sqrt{(x^{(1)})^2 + (x^{(2)})^2}} & -\frac{x^{(1)}}{\sqrt{(x^{(1)})^2 + (x^{(2)})^2}} \end{bmatrix}. \quad (\text{A.2})$$

Numerical solutions to Equation A.1 can be seen in the left-hand panel of Figure A.1. One can imagine moving along any of the plotted δ_1 -integral curves — for any \mathcal{S} -transformation based on orthogonal Jacobian matrices, this corresponds to the first transformed state variable changing, while the second stays constant. Movement along a δ_2 -integral curve corresponds to changes in the second \mathcal{S} -transformed state variable, which, because $\delta_2(\cdot)$ gives an unspanned direction (recall that $\delta_2(x)$ is almost-everywhere a basis for \mathcal{S}^x), has no effect on the term structure.

There would be many ways to suitably index the points along these integral curves. For example, for the δ_2 -integral curves, one can consider a particular curve and use the parameter that results

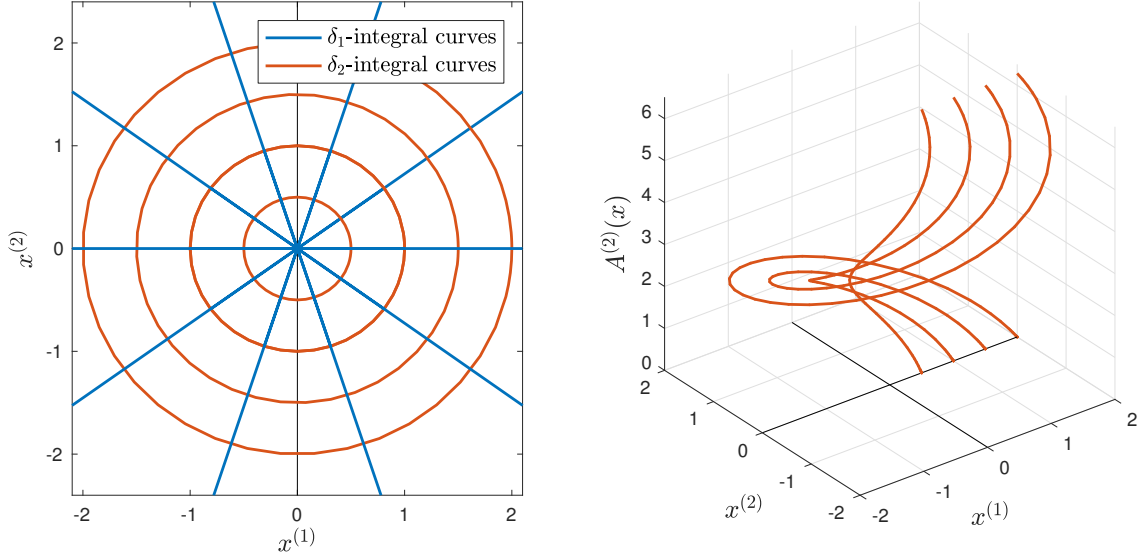


Figure A.1: The left-hand panel shows numerical solutions to Equation A.1 for the two vector fields given by Equation A.2, for various initial values. These are the integrals curves of either vector field, and can also be thought of as coordinate curves. The right-hand panel plots the proposed $A^{(2)}(x)$ for all x along the δ_2 -integral curves plotted in the left-hand panel. The values taken on by $A^{(2)}(x)$ are based on the parameter values of one of these curves.

from Equation A.1 as the index: let $\{h_*^{(2)}(r)\}$ be the specific solution to Equation A.1 based on $h_*^{(2)}(0) = [1 \ 0]^\top$. One can then restrict $h_*^{(2)}(\cdot)$ to only take arguments in $[0, 2\pi)$ and still obtain the same image. A particular argument $r_* \in [0, 2\pi)$ unambiguously describes a position on the curve $\{h_*^{(2)}(r)\}$, and in fact on any other δ_2 -integral curve — one can simply connect other points to the reference curve $\{h_*^{(2)}(r)\}$ with a δ_1 -integral curve, which is orthogonal to $\delta_2(\cdot)$ and therefore does not affect the second coordinate/state variable. Then one sets $A^{(2)}(x)$ by firstly corresponding x to a point along $\{h_*^{(2)}(r)\}$, and then taking the parameter that corresponds to this point. This gives a viable $A^{(2)}(\cdot)$. A restricted version of this $A^{(2)}(\cdot)$ is plotted in right-hand panel of Figure A.1.³

The restricting of the domain wherever the integral curves intersect is important for two reasons. Firstly, this effectively restricts the domain of the inverse transformation $\hat{\mathcal{X}}$, which is instrumental in ensuring invertibility — further comments are given below. Secondly, it ensures that $\hat{\mathcal{X}}$ will be given by a product of real-valued sets (either the real numbers themselves, or intervals thereof), which in turn ensures the convexity of $\hat{\mathcal{X}}$ (see the discussion in Footnote 7 in Section 3.1).

This method results in almost-everywhere differentiable transformation functions. This can best be seen by considering the inverse transformation, which takes a parameter that indicates a position on the reference integral curve. This is then used as an initial condition for an orthogonal integral curve, and these integral curves depend smoothly on their initial conditions [Lee, 2002, Ch.17]. Excepting the $X_*(\lambda \otimes \mathbb{P})$ -null states outside of $\mathcal{X}_S^{(\bar{p})}$, the only non-differentiable points correspond to any endpoints of the curve domains (this can be seen clearly seen on right-hand panel of Figure A.1) — if an δ_i -integral curves extend indefinitely, $A^{(i)}(\cdot)$ is differentiable everywhere in $\mathcal{X}_S^{(\bar{p})}$. Using this method to suggest $A^{(1)}(\cdot)$ would not, however, result in a \mathcal{C}^2 function, which is required by Definition 3.1.14. The strategy we suggest below addresses this.

³The proposed $A^{(2)}(\cdot)$ is very similar to the $\text{atan2}(\cdot, \cdot)$ function used in Equation 3.16, which indexes the points along the δ_2 -integral curves in a way that differs only by straightforward translations. In a case with $n > 2$, there would be more variation among the viable \mathcal{S} -transformations, as there would be more freedom in selecting orthonormal vector fields.

However the points along the integral curves are grouped and indexed, the transformation's local invertibility — which it features from the orthonormal (and thus invertible) Jacobian — should be extendable to the whole domain. To see this, note firstly that for any i , we have $\delta_i(x) \neq \mathbf{0}_n$ for all $x \in \mathcal{X}_S^{(\bar{p})}$; this ensures that the integral curves do in fact extend over the state space, so that $A^{(i)}(\cdot)$ is in fact a function on the domain $\mathcal{X}_S^{(\bar{p})}$ (if an integral curve had only one point in its image, all of its index values would correspond to a single state x). The points outside of $\mathcal{X}_S^{(\bar{p})}$ should be manually assigned to preserve invertibility (see, for example, Equation 3.16). Secondly, the ability to restrict the domain of the integral curves, and therefore the transformed state space $\hat{\mathcal{X}}$ ensures that $A(\cdot)$ is surjective. Finally, any suitable indexing would preclude the same index value being assigned to different positions along the integrals, and so $A(\cdot)$ is straightforwardly injective (this is related to the first point: if an integral curve's image is a single state, the injectiveness of $A(\cdot)$ would be compromised).

The only requirement of the definition of an \mathcal{S} -transformation (Definition 3.1.14) that has not been addressed is the \mathcal{C}^2 requirement for the first \bar{m} components of $A(\cdot)$. We conclude the appendix by proposing, for $i = 1, \dots, \bar{m}$, that one sets

$$A^{(i)}(x) = P(\tau_i, x), \quad (\text{A.3})$$

for some $\tau_i > 0$. In practice, one is able to find \bar{m} maturities $\{\tau_i\}$ such that the \bar{m} vectors $\nabla A^{(i)}(x) = \nabla P(\tau_i, x)$ are independent, and therefore form a basis for $(\mathcal{S}^x)^\perp$.⁴ Though not orthogonal, defining the first \bar{m} state variables in this way is very convenient — the \mathcal{C}^2 requirement of an \mathcal{S} -transformation is met (from Assumption 2.1.2), and the explicit nature of the transformation allows one to compute the dynamics of spanned component (see Lemma 3.1.18).⁵

Using the bond pricing function to determine the first \bar{m} state variables, the integral-curve method can then be used to determine functions to accompany them. One must select \bar{p} orthonormal vector fields that $X_*(\lambda \otimes \mathbb{P})$ -almost everywhere form a basis for \mathcal{S}^x , then propagate their integral curves with Equation A.1, and seek or construct functions that increase along these curves (and thus index the state's position relative to the curves) but are level in orthogonal directions. Provided the ODEs admit solutions unproblematically, and the local invertibility of the transformation can be extended (which may require a manual assignment of points outside $\mathcal{X}_S^{(\bar{p})}$), the requirements of an \mathcal{S} -transformation can be met in this way.

⁴Corollary 3.3.7 effectively states that this is indeed possible for any particular state x (outside of $\bigcup_{j=\bar{p}+1}^n \mathcal{X}_S^{(j)}$). Assumption 3.3.8 conveys the fact that, typically, this is not only possible, but is the case for *any* set of distinct maturities. Our assumption here is stronger than the implication of Corollary 3.3.7 but weaker than Assumption 3.3.8 — we require one set of maturities \bar{m} , but they must give rise to \bar{m} vectors $\nabla P(\tau_i, x)$ that form a basis for $(\mathcal{S}^x)^\perp$.

⁵Note that these dynamics are not the ZCB dynamics given by Equation 2.5, because here we are holding *time to maturity*, rather than maturity time, constant (these state variables do not therefore represent the prices of tradeable assets).

Appendix B

Invertible Matrix Processes

We first address the matrix process from Section 3.3.1, after which we extend the remarks to cover Section 3.3.3.

In order to relax Assumption 3.3.2, we need a specific assumption to address the first potential problem described in the first subsection of Section 3.3.1 (namely, that the inability to find suitable maturities when $X_t \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}$ might cause invertibility problems *around* this set, not merely inside it).

Assumption B.0.1. If $X_t \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}$, we assume that there exist d maturity times $\mathcal{T}^* = \{T_i\}$ (all greater than t), giving rise to a constant maturity process $\{\mathcal{T}_t^*\}$ such that $\nabla P_t^{\mathcal{T}^*} \sigma(X_t)$ is almost surely invertible over an open interval following t ; that is, $\inf\{\epsilon \geq 0 \mid \det(\nabla P_{t+\epsilon}^{\mathcal{T}^*} \sigma(X_{t+\epsilon})) \neq 0\} = 0$ \mathbb{P} -almost surely.

This is more a technical than a conceptual requirement of a model — there is no economic or practical relevance in the ability to preselect maturities in this way, given that $\bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)} X_*(\lambda \otimes \mathbb{P})$ -null. The important conceptual feature is whether or not *general maturity availability* (Assumption 3.3.2) prevails.

Consider then applying Lemma 3.3.1 algorithmically, starting at $t = 0$, in the following sense. Provided that $X_0 \in \mathcal{X}_{\mathcal{U}}^{(0)}$, by Lemma 3.3.1 one can find d suitable maturity times \mathcal{T}_1 that result in invertibility. If $X_0 \notin \mathcal{X}_{\mathcal{U}}^{(0)}$, we must appeal to Assumption B.0.1 for maturities \mathcal{T}_1 that result in invertibility immediately after $t = 0$. The maturities \mathcal{T}_1 can then be maintained for some finite period, until (or before) $\nabla P_t^{\mathcal{T}_1} \sigma(X_t)$ becomes non-invertible or one of the maturities is reached, at which point the algorithm can be re-applied — d new maturities \mathcal{T}_2 are identified as they were at $t = 0$. Note that the continuity of the determinant of $\{\nabla P_t^{\mathcal{T}_i} \sigma(X_t)\}$ ensures that invertibility is indeed preserved for some interval.¹ These become the second constant values that the maturity process $\{\mathcal{T}_t\}$ takes on, and are maintained for some (λ -non-null) interval of time. This process can be continued until the finite time horizon S is reached. While $\{\nabla P_t^{\mathcal{T}} \sigma(X_t)\}$ is not necessarily invertible at $t = 0$ or other time points when new maturities are engaged, these points have a Lebesgue measure zero, and so the matrix process $\{\nabla P_t^{\mathcal{T}} \sigma(X_t)\}$ is $\lambda \otimes \mathbb{P}$ -almost surely invertible.

Finally, consider the possibility of this algorithm converging before the time horizon S is exceeded (i.e., the lengths of the intervals of invertibility having a sum that does not exceed S). Of course, one could induce this oneself by selecting ZCBs with increasingly near maturities, but suppose that one cannot make a maturity selection that avoids convergence. If the algorithm

¹The determinant is a continuous function of, in this case, a continuous matrix process (the continuity of the diffusion $\{X_t\}$, the assumed continuous differentiability of the bond pricing function $P(\cdot, \cdot)$ and the assumed continuity of the diffusion function $\sigma(\cdot)$ together imply the continuity of $\{\nabla P_t^{\mathcal{T}_i} \sigma(X_t)\}$).

converged to some $S^* \leq S$, the matrix process is still $\lambda \otimes \mathbb{P}$ -almost surely invertible over $[0, S^*]$, as invertibility can only be absent at the countably infinite time points where the maturities are reset. At S^* one could then appeal to Lemma 3.3.1 (or Assumption B.0.1, if $X_{S^*} \in \bigcup_{j=1}^d \mathcal{X}_{\mathcal{U}}^{(j)}$) and resume the algorithm — this resembles the limit case in transfinite induction — eventually resulting in $\lambda \otimes \mathbb{P}$ -almost sure invertibility.

In order to relax Assumption 3.3.8 in Section 3.3.3, we require an analogue of Assumption B.0.1.

Assumption B.0.2. If $\hat{X}_t \in \bigcup_{j=p+1}^d \hat{\mathcal{X}}_{\mathcal{U}}^{(j)}$, we assume that there exists m maturity times $\mathcal{T}^* = \{T_i\}$ (all greater than t), giving rise to a constant maturity process $\{\mathcal{T}_t^*\}$, such that $\nabla \tilde{P}_t^{\mathcal{T}^*} \sigma^Y(\hat{X}_t)$ is almost surely invertible over an open interval following t ; that is, $\inf\{\epsilon \geq 0 \mid \det(\nabla \tilde{P}_{t+\epsilon}^{\mathcal{T}^*} \sigma^Y(\hat{X}_{t+\epsilon})) \neq 0\} = 0$ \mathbb{P} -almost surely. Similarly, if $\hat{X}_t \in \bigcup_{j=\bar{p}+1}^n \hat{\mathcal{X}}_{\mathcal{S}}^{(j)}$, we assume that there exists \bar{m} maturity times $\bar{\mathcal{T}}^* = \{\bar{T}_i\}$ (all greater than t), giving rise to a constant maturity process $\{\bar{\mathcal{T}}_t^*\}$, such that $\nabla \tilde{P}_t^{\bar{\mathcal{T}}^*}$ is almost surely invertible over an open interval following t ; that is, $\inf\{\epsilon \geq 0 \mid \det(\nabla \tilde{P}_{t+\epsilon}^{\bar{\mathcal{T}}^*}) \neq 0\} = 0$ \mathbb{P} -almost surely.

An algorithmic application of Corollaries 3.3.6 and 3.3.7, in the sense described above, is then possible.

Appendix C

Partial Derivatives of the Bond Pricing Function

C.1 The Short Rate and the Bond Pricing Function

In Equation 2.2, the short rate function $r : [0, S] \times \mathcal{X}$ is defined with

$$r(t, x) = - \frac{\partial \log P(t, T, x)}{\partial T} \Big|_{T=t}. \quad (\text{C.1})$$

Using the chain rule, one gets

$$\begin{aligned} r(t, x) &= - \frac{\frac{\partial P(t, T, x)}{\partial T}}{P(t, T, x)} \Big|_{T=t} \\ &= - \frac{\frac{\partial P(t, T, x)}{\partial T} \Big|_{T=t}}{P(t, t, x)} \\ &= - \frac{\partial P(t, T, x)}{\partial T} \Big|_{T=t} \\ &= -P_T(t, t, x), \end{aligned} \quad (\text{C.2})$$

where the final line uses the partial-derivative notation of Chapter 5. Because the short rate is given by the first state variable, we conclude that $P_T(t, t, x) = -x^{(1)}$. When Definition 5.0.4 holds, we have that $P_T(t, t, r) = -r$.

From a mathematical perspective, Equation C.2 is a valid alternative to Equation C.1 because the natural-logarithm function at an argument of value one (which the bond takes on at $T = t$) preserves the slope of any function it is composed with. Financially, Equation C.2 expresses the return from investing in a short-term bond, at the limit of the term tending to zero:

$$\begin{aligned} \lim_{\Delta t \downarrow 0} \frac{P(T, T, x) - P(T - \Delta t, T, x)}{\Delta t} &= \lim_{\Delta t \downarrow 0} \frac{P(T - \Delta t, T - \Delta t, x) - P(T - \Delta t, T, x)}{\Delta t} \\ &= - \frac{\partial P(t, T, x)}{\partial T} \Big|_{T=t}, \end{aligned}$$

whereas Equation C.1 expresses the continuously compounded return of the same investment:

$$\begin{aligned}
\lim_{\Delta t \downarrow 0} \frac{\log \left(\frac{P(T, T, x)}{P(T - \Delta t, T, x)} \right)}{\Delta t} &= \lim_{\Delta t \downarrow 0} \frac{\log P(T, T, x) - \log P(T - \Delta t, T, x)}{\Delta t} \\
&= \lim_{\Delta t \downarrow 0} \frac{\log P(T - \Delta t, T - \Delta t, x) - \log P(T - \Delta t, T, x)}{\Delta t} \\
&= - \frac{\partial \log P(t, T, x)}{\partial T} \Big|_{T=t}.
\end{aligned}$$

The return (annualised but compounded once) converges to the (annualised) continuously compounded return as the investment period becomes short.

C.2 Sensitivity to the Short Rate

Recall that the expression in Lemma 5.1.1 contains $P_r(t, T, r)$ in the denominator, which is this required to be non-zero.

Consider differentiating the identity $P_T(t, t, r) = -r$ with respect to the short rate:

$$P_{rT}(t, t, r) = -1.$$

Due to the twice continuous differentiability from Assumption 5.0.5, first-order partial derivatives are symmetrical:

$$P_{Tr}(t, t, r) = -1. \tag{C.3}$$

Noting that there is no short-rate sensitivity at maturity ($P_r(t, t, r) = \frac{\partial}{\partial r} 1 = 0$), Equation C.3 guarantees that a non-zero (negative) short-rate sensitivity is attained as soon any finite maturity time is given to the bond in question ($P_r(t, t + \Delta t, r) \approx -\Delta t$) — if short-rate sensitivity remained zero near maturity, we would have $P_{Tr}(t, t, r) = 0$, contradicting Equation C.3.

This is the place where the assumption of twice *continuous* differentiability in maturity time is needed. An alternative approach would be to relax this to twice differentiability, and assume the natural economic condition that

$$P_r(t, T, r) \neq 0,$$

so that the ratio in Lemma 5.1.1 is well-defined in general.

Appendix D

Comparison to Linear-rational Models

D.1 (1,1) Models in the Linear-rational Framework

In a bivariate, Brownian-based linear-rational model, the two-dimensional state process $\{X_t\}$ satisfies the following stochastic differential equation under the risk-neutral measure:¹

$$dX_t = \left(\kappa(\theta - X_t) + \frac{\sigma(X_t)\sigma(X_t)^\top \mathbf{1}}{1 + \mathbf{1}^\top X_t} \right) dt + \sigma(X_t) dW_t, \quad (\text{D.1})$$

where $\kappa \in \mathbb{R}^{2 \times 2}$, $\theta \in \mathbb{R}^2$, $\mathbf{1} = [1 \ 1]^\top$, $\sigma(\cdot)$ is a suitable volatility coefficient function (so that $\sigma : \mathbb{R}^2 \rightarrow \mathbb{R}^{2 \times 2}$) and where $\{W_t\}$ is a two-dimensional Brownian motion under \mathbb{Q} . The short rate is related to the state variables with²

$$r_t = \alpha - \frac{\mathbf{1}^\top \kappa(\theta - X_t)}{1 + \mathbf{1}^\top X_t},$$

where α is a constant. In order for one of the state variables to be unspanned, κ must be set to a multiple of the identity matrix.³ Suppose the diagonal entries of κ are equal to k . It is convenient to define a new state variable with $Z_t = \mathbf{1}^\top X_t$, which alone can describe the short rate:

$$r_t = \alpha - \frac{k(\theta_Z - Z_t)}{1 + Z_t}, \quad (\text{D.2})$$

¹Filipović et al. [2017] specify their pricing kernel process as per Equation 4.1: $\{e^{-\alpha t}(\varphi + \psi^\top X_t)\}$, but later specialise this to $\{e^{-\alpha t}(1 + \mathbf{1}^\top X_t)\}$. Provided one is not interested in setting $\varphi = 0$, no generality is lost by doing so. We work with this more parsimonious version.

²The pricing kernel process in Footnote 1 implies a rational bond pricing function as per Equation 1.11. Via Equation 2.2, this gives an expression for the short rate.

³Strictly speaking, one could also set

$$\kappa = \begin{bmatrix} k_1 & k_2 \\ k_2 & k_1 \end{bmatrix},$$

but, like the issue in Footnote 1, one can ignore these cases without sacrificing generality, as they are captured by certain specifications with diagonal κ .

where $\theta_Z = \mathbf{1}^\top \theta$. The dynamics of the short rate are then implied by Itô's lemma:⁴

$$dr_t = \left(\frac{k^2(\theta_Z - Z_t)(1 + \theta_Z)}{(1 + Z_t)^2} \right) dt + \frac{k(1 + \theta_Z)}{(1 + Z_t)^2} \mathbf{1}^\top \sigma(X_t) dW_t,$$

which can be simplified, using Equation D.2, to

$$dr_t = (r_t - \alpha)(r_t - \alpha - k) dt + \frac{(r_t - \alpha - k)^2}{k(1 + \theta_Z)} \mathbf{1}^\top \sigma(X_t) dW_t. \quad (\text{D.3})$$

Note, firstly, that the final part of the volatility is left in terms of the original state variables — to finalise the transformation, one would need to define a second state variable $\{u_t\}$ so that the volatility can be described in terms of the new state variables: $\sigma(X_t) = \hat{\sigma}(r_t, u_t)$. Secondly, USV is not necessarily exhibited: depending on how $\sigma(\cdot)$ is specified, the short rate's total volatility may or may not depend on the second state variable (a condition that we check with Definition 5.0.4). Thirdly, the Brownian motion can be transformed (with an orthonormal transformation described by Definition 2.2.8) so that Equation D.3 conforms to the lower triangularity in Assumption 5.0.3. These three points are illustrated for a specific case in Appendix D.2 below.

Finally, and most importantly, we can now see how linear-rational models encompass a wide class of (1,1) models. It follows from Theorem 5.1.3 that (1,1) models are given by any specification of the dynamics in Assumption 5.0.3, provided that $\mu^{\mathbb{Q},r}(t, r, u) = r^2 + r\alpha_1(t) + \alpha_2(t)$. Ignoring the possibility of time inhomogeneity, Equation D.3 shows that bivariate linear-rational models attain all of the permitted short-rate drift functions.⁵ Using the freedom to specify the original volatility function $\sigma(\cdot)$, they allow any (time-homogeneous) specification of the volatility matrix in Assumption 5.0.3 to be achieved.⁶ For example, one could set

$$\sigma(X_t) = \begin{bmatrix} \frac{k(1+\theta_Z)\sigma_A(X_t)}{(r_t-\alpha-k)^2} - \sigma_B(X_t) & -\sigma_C(X_t) \\ \sigma_B(X_t) & \sigma_C(X_t) \end{bmatrix}$$

for scalar valued functions $\sigma_A(\cdot)$, $\sigma_B(\cdot)$ and $\sigma_C(\cdot)$. Defining $u_t = X_t^{(2)}$, this implies

$$d \begin{bmatrix} r_t \\ u_t \end{bmatrix} = \begin{bmatrix} (r_t - \alpha)(r_t - \alpha - k) \\ k(\theta^{(2)} - u_t) - \frac{\sigma_A(X_t)\sigma_B(X_t)}{r_t - \alpha - k} + \frac{\sigma_B(X_t)^2(r_t - \alpha - k)}{k(1 + \theta_Z)} \end{bmatrix} dt + \begin{bmatrix} \sigma_A(X_t) & 0 \\ \sigma_B(X_t) & \sigma_C(X_t) \end{bmatrix} dW_t,$$

where $\theta_t^{(2)} = [0 \ 1] \theta$. One can then choose $\sigma_A(\cdot)$, $\sigma_B(\cdot)$ and $\sigma_C(\cdot)$ so that, when converted to be in terms of r_t and u_t , they give any desired (time-homogeneous) volatility functions.⁷ However, the form of $\mu^{\mathbb{Q},u}(\cdot, \cdot, \cdot)$ — the risk-neutral drift of the unspanned volatility process — is a result of any such specification and cannot be separately controlled.

⁴The short rate is a function of Z_t , as per Equation D.2. Letting $f(\cdot)$ denote this function, one finds that

$$\frac{\partial f(z)}{\partial z} = \frac{k(1 + \theta_Z)}{(1 + z)^2} \quad \text{and} \quad \frac{\partial^2 f(z)}{\partial z^2} = \frac{-2k(1 + \theta_Z)}{(1 + z)^3}.$$

Itô's lemma, based on the dynamics of $\{Z_t\}$ implicit in Equation D.1, can then be applied. The second-order Itô term cancels exactly with the contribution of the non-linear part of the drift in Equation D.1.

⁵The cases where the quadratic $\mu^{\mathbb{Q},r}(\cdot, \cdot, \cdot)$ has no real roots are irrelevant, as these cannot be made to satisfy the linear growth condition that prevents explosions.

⁶Though Filipović et al. [2017] do not consider time-inhomogeneous diffusions, a time-inhomogeneous volatility would be a trivial extension. Their bond pricing result does, however, rely on a time-inhomogeneous drift specification, while our bond pricing results do not.

⁷For example, to achieve the short-rate volatility function used in the LADQ(1,1) model, one must set

$$\sigma_A(X_t) = \left(\alpha - \frac{k(\theta_Z - \mathbf{1}^\top X_t)}{1 + \mathbf{1}^\top X_t} \right) \left(\bar{r} - \alpha + \frac{k(\theta_Z - \mathbf{1}^\top X_t)}{1 + \mathbf{1}^\top X_t} \right) X_t^{(2)},$$

as this gives $\sigma_A(X_t) = \sigma^r(r_t, u_t) = r_t(\bar{r} - r_t)u_t$ (using Equation D.3 and maintaining that $u_t = X_t^{(2)}$).

The class of linear-rational models, therefore, includes all time-homogeneous (1,1) models that have a certain risk-neutral drift for their unspanned volatility process. In other words, when constructing a time-homogeneous (1,1) model, one can remain in the linear-rational framework and retain freedom over all model aspects except the choice of $\mu^{\mathbb{Q},u}(\cdot, \cdot, \cdot)$. Note that a different definition of $\{u_t\}$ does not ameliorate this restriction — different forms of $\mu^{\mathbb{Q},u}(\cdot, \cdot, \cdot)$ would result, but not forms that can be freely chosen for a given specification of the volatility $\sigma(\cdot)$.

This accords with the intuition that linear-rational models allow any volatility specification, but require a specific drift form (for a given volatility). Strictly speaking, this point extends to the drift under the real-world measure; however, Filipović et al. [2017] embellish their models — by introducing an auxiliary probability measure — so that real-world drift functions can be generalised.

Equation D.3 shows the short rate in bivariate linear-rational models to be bounded above:

$$\sup_{z \in (-1, \infty)} \alpha - \frac{k(\theta_Z - z)}{1 + z} = \lim_{z \rightarrow \infty} \alpha - \frac{k(\theta_Z - z)}{1 + z} = \alpha + k.$$

The supremum is taken over the widest possible range for $\{Z_t\}$ that maintains positivity of the pricing kernel process. Note that $\alpha + k$ corresponds to our λ_2 — linear-rational models thus endogenously avoid the possible explosions of the short rate discussed towards the end of Section 5.1.

Filipović et al. [2017] show that ZCB prices in these linear-rational models are given by

$$\begin{aligned} P_{tT} &= e^{-\alpha(T-t)} \frac{1 + \mathbf{1}^\top \theta + \mathbf{1}^\top e^{-\kappa(T-t)} (X_t - \theta)}{1 + \mathbf{1}^\top X_t} \\ &= e^{-\alpha(T-t)} \frac{1 + \theta_Z + e^{-k(T-t)} (Z_t - \theta_Z)}{1 + Z_t}, \end{aligned}$$

which, via Equation D.3, simplifies to an affine function in the short rate:

$$P_{tT} = \frac{e^{-\alpha(T-t)}}{k} \left((r_t - \alpha)(e^{-k(T-t)} - 1) + k \right),$$

as required by Theorem 5.1.3 (one can also verify that this agrees, after a suitable parameter translation, with the specific pricing function from Corollary 5.1.5).

D.2 The Short Rate in the LRSQ(1,1) Model

We now specialise Appendix D.1 — where a bivariate linear-rational diffusion model is cast in terms of the short rate and risk-neutral measure — to the LRSQ(1,1) model, where the state variable volatility in Equation D.1 is given by

$$\sigma(X_t) = \text{diag} \left(\sigma_i \sqrt{X_t^{(i)}} \right) \sigma(X_t) = \begin{bmatrix} \sigma_1 \sqrt{X_t^{(1)}} & 0 \\ 0 & \sigma_2 \sqrt{X_t^{(2)}} \end{bmatrix},$$

where $\sigma_1, \sigma_2 > 0$. Defining $u_t = X_t^{(2)}$, it follows — from Equations D.2 and D.3 — that in the LRSQ(1,1) model, the short rate satisfies

$$dr_t = (r_t - \alpha)(r_t - \alpha - k) dt + \frac{(r_t - \alpha - k)^2}{k(1 + \theta_Z)} \left[\sigma_1 \sqrt{\frac{\alpha - k\theta_Z - r_t}{r_t - \alpha - k}} - u_t \quad \sigma_2 \sqrt{u_t} \right] dW_t.$$

It is possible to define the second state variable in different ways, which would recast, but not fundamentally change the model. This definition aligns with the transformation recommended by Filipović et al. [2017], and will make the contribution of the unspanned volatility transparent and amenable. Then, from Equation D.1 (in the context of the LRSQ(1,1) model) and Equation D.2, we have

$$du_t = \left(k(\theta^{(2)} - u_t) - \frac{u_t \sigma_2^2 (r_t - \alpha - k)}{k(1 + \theta_Z)} \right) dt + \sigma_2 \sqrt{u_t} dW_t^{(2)}.$$

An orthonormal transformation can be used to define a new Brownian motion $\{\hat{W}_t\}$ — with the orthonormal basis chosen as described in the beginning of Section 3.3.3 — such that the specification gives the lower triangularity in Assumption 5.0.3:

$$dr_t = (r_t - \alpha)(r_t - \alpha - k) dt + \frac{(r_t - \alpha - k)^2}{k(1 + \theta_Z)} \sqrt{\frac{\sigma_1^2 (\alpha - k\theta_Z - r_t)}{r_t - \alpha - k} + u_t (\sigma_2^2 - \sigma_1^2)} d\hat{W}_t^{(1)}.$$

It is clear that USV will result if and only if $\sigma_2^2 \neq \sigma_1^2$ — these are the cases in which $\{u_t\}$ has an effect on short-rate volatility (causing Definition 5.0.4 to be satisfied). The dynamics of $\{u_t\}$ are now given by

$$du_t = \left(k(\theta^{(2)} - u_t) - \frac{u_t \sigma_2^2 (r_t - \alpha - k)}{k(1 + \theta_Z)} \right) dt + \sigma_2 \sqrt{u_t} \left[\frac{\sigma_2 \sqrt{u_t}}{\sqrt{\frac{\sigma_1^2 (\alpha - k\theta_Z - r_t)}{r_t - \alpha - k} + u_t (\sigma_2^2 - \sigma_1^2)}} \sqrt{1 - \frac{\sigma_2^2 u_t}{\frac{\sigma_1^2 (\alpha - k\theta_Z - r_t)}{r_t - \alpha - k} + u_t (\sigma_2^2 - \sigma_1^2)}} \right] d\hat{W}_t.$$

After a parameter translation (namely, $\alpha = \lambda_1$ and $k = \lambda_2 - \lambda_1$) and definition of a correlation, this leads the LRSQ(1,1) short-rate dynamics in Equations 5.23 and 5.24 in Section 5.2.2.

Like in the LADQ(1,1) model, the LRSQ(1,1) short rate is bounded. Its upper bound is discussed above in Appendix D.1. Its lower bound follows from $\{Z_t\}$, which is the sum of square-root processes (either under \mathbb{P} or the auxiliary measure mentioned above), being non-negative. From Equation D.2, one sees that the short rate is an increasing function of $\{Z_t\}$ and is thus bounded below by $\alpha - k\theta_Z$ (Filipović et al. [2017] suggest setting $\alpha = k\theta_Z$ to maintain a non-negative short rate).

Bibliography

- Ahn, D. H., Dittmar, R. F., & Gallant, A. R. (2002). Quadratic term structure models: theory and evidence. *Review of Financial Studies*, 15(1), 243–288.
- Ahn, D. H. & Gao, B. (1999). A parametric nonlinear model of term structure dynamics. *Review of Financial Studies*, 12(4), 721–762.
- Andersen, L. (2011). Option pricing with quadratic volatility: a revisit. *Finance and Stochastics*, 15(2), 191–219.
- Andersen, T. G. & Benzoni, L. (2010). Do bonds span volatility risk in the us treasury market? A specification test for affine term structure models. *Journal of Finance*, 65(2), 603–653.
- Bianchetti, M. & Carlicchi, M. (2011). Interest rates after the credit crunch: Multiple curve vanilla derivatives and SABR. Working paper, SSRN 1783070.
- Bikbov, R. & Chernov, M. (2009). Unspanned stochastic volatility in affine models: evidence from eurodollar futures and options. *Management Science*, 55(8), 1292–1305.
- Björk, T. (2004). *Arbitrage theory in continuous time*. Oxford University Press.
- Björk, T. & Svensson, L. (2001). On the existence of finite-dimensional realizations for nonlinear forward rate models. *Mathematical Finance*, 11(2), 205–243.
- Brigo, D. & Mercurio, F. (2001). A deterministic-shift extension of analytically-tractable and time-homogeneous short-rate models. *Finance and Stochastics*, 5(3), 369–387.
- Brigo, D. & Mercurio, F. (2007). *Interest rate models - theory and practice: with smile, inflation and credit*. Springer.
- Carmona, R. & Tehranchi, M. R. (2007). *Interest rate models: an infinite dimensional stochastic analysis perspective*. Springer Science & Business Media.
- Carr, P., Gabaix, X., & Wu, L. (2011). Linearity-generating processes, unspanned stochastic volatility, and interest-rate option pricing. Working paper, New York University.
- Casassus, J., Collin-Dufresne, P., & Goldstein, B. (2005). Unspanned stochastic volatility and fixed income derivatives pricing. *Journal of Banking and Finance*, 29(11), 2723–2749.
- Cheridito, P. & Gabaix, X. (2008). Regularity conditions to ensure the existence of linearity-generating processes. Working paper, Princeton University.
- Chiarella, C. & Kwon, O. K. (2003). Finite dimensional affine realisations of HJM models in terms of forward rates and yields. *Review of Derivatives Research*, 6(2), 129–155.

- Collin-Dufresne, P. & Goldstein, R. (2001). Stochastic correlation and the relative pricing of caps and swaptions in a generalized-affine framework. Working paper, SSRN 286322.
- Collin-Dufresne, P. & Goldstein, R. S. (2002). Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. *Journal of Finance*, 57(4), 1685–1730.
- Collin-Dufresne, P., Goldstein, R. S., & Jones, C. S. (2009). Can interest rate volatility be extracted from the cross section of bond yields? *Journal of Financial Economics*, 94(1), 47–66.
- Constantinides, G. M. (1992). A theory of the nominal term structure of interest rates. *Review of Financial Studies*, 5(4), 531–552.
- Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2), 385–407.
- Crépey, S. (2013). *Financial modeling*. Springer.
- Dai, Q. & Singleton, K. J. (2000). Specification analysis of affine term structure models. *Journal of Finance*, 55(5), 1943–1978.
- Dana, R.-A. & Jeanblanc, M. (2007). *Financial markets in continuous time*. Springer Science & Business Media.
- Delbaen, F. & Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen*, 300(1), 463–520.
- Derman, E. & Kani, I. (1994). Riding on a smile. *Risk*, 7(2), 32–39.
- Dothan, L. U. (1978). On the term structure of interest rates. *Journal of Financial Economics*, 6(1), 59–69.
- Driessen, J., Klaassen, P., & Melenberg, B. (2003). The performance of multi-factor term structure models for pricing and hedging caps and swaptions. *Journal of Financial and Quantitative Analysis*, 38(03), 635–672.
- Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance*, 57(1), 405–443.
- Duffee, G. R. (2011). Information in (and not in) the term structure. *Review of Financial Studies*, 24(9), 2895–2934.
- Duffee, G. R. & Stanton, R. H. (2012). Estimation of dynamic term structure models. *The Quarterly Journal of Finance*, 2(02), 1–51.
- Duffie, D. & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6(4), 379–406.
- Dupire (1994). Pricing with a smile. *Risk*, 7(1), 18–20.
- Engle, R. F. & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2), 251–276.
- Fan, R., Gupta, A., & Ritchken, P. (2003). Hedging in the possible presence of unspanned stochastic volatility: Evidence from swaption markets. *Journal of Finance*, 58(5), 2219–2248.

- Fan, R., Gupta, A., & Ritchken, P. (2007). On pricing and hedging in the swaption market: How many factors, really? *Journal of Derivatives*, 15(1), 9–33.
- Feller, W. (1951). Two singular diffusion problems. *Annals of Mathematics*, 54, 173–182.
- Filipović, D. (2009). *Term-structure models: a graduate course*. Springer.
- Filipović, D., Larsson, M., & Statti, F. (2017). Unspanned stochastic volatility in the multi-factor CIR model. Working paper, SSRN 2964751.
- Filipović, D., Larsson, M., & Trolle, A. B. (2017). Linear-rational term structure models. *Journal of Finance*, 72(2), 655–704.
- Fong, H. G. & Vasiček, O. A. (1991). Fixed-income volatility management. *Journal of Portfolio Management*, 17(4), 41–46.
- Gabaix, X. (2009). Linearity-generating processes: A modelling tool yielding closed forms for asset prices. Working paper, New York University.
- Goldstein, R. S. (2000). The term structure of interest rates as a random field. *Review of Financial Studies*, 13(2), 365–384.
- Golub, G. H. & Van Loan, C. F. (2012). *Matrix computations*. JHU Press.
- Grbac, Z. & Runggaldier, W. J. (2015). *Interest rate modeling: post-crisis challenges and approaches*. Springer.
- Gupta, A. & Subrahmanyam, M. G. (2005). Pricing and hedging interest rate options: Evidence from cap–floor markets. *Journal of Banking and Finance*, 29(3), 701–733.
- Hagan, P. S., Kumar, D., Lesniewski, A. S., & Woodward, D. E. (2002). Managing smile risk. *Wilmott Magazine*, 1, 84–108.
- Han, B. (2007). Stochastic volatilities and correlations of bond yields. *Journal of Finance*, 62(3), 1491–1524.
- Harrison, J. M. & Kreps, D. M. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20(3), 381–408.
- Harrison, J. M. & Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications*, 11(3), 215–260.
- Haubrich, J., Pennacchi, G., & Ritchken, P. (2012). Inflation expectations, real rates, and risk premia: evidence from inflation swaps. *Review of Financial Studies*, 25(5), 1588–1629.
- Heath, D., Jarrow, R., & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60(1), 77–105.
- Heidari, M. & Wu, L. (2003). Are interest rate derivatives spanned by the term structure of interest rates? *Journal of Fixed Income*, 13(1), 75–86.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2), 327–343.
- Hull, J. & White, A. (1990). Pricing interest-rate-derivative securities. *Review of Financial Studies*, 3(4), 573–592.

- Ingersoll, J. E. (1997). Valuing foreign exchange rate derivatives with a bounded exchange process. *Review of Derivatives Research*, 1(2), 159–181.
- Jamshidian, F. (1989). An exact bond option formula. *Journal of Finance*, 44(1), 205–209.
- Jarrow, R., Li, H., & Zhao, F. (2007). Interest rate caps “smile” too! But can the LIBOR market models capture the smile? *Journal of Finance*, 62(1), 345–382.
- John, F. (1982). *Partial differential equations*. Springer.
- Joslin, S. (2017). Can unspanned stochastic volatility models explain the cross section of bond volatilities? *Management Science*, (forthcoming).
- Joslin, S., Priebisch, M., & Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *Journal of Finance*, 69(3), 1197–1233.
- Jost, J. (2011). *Riemannian geometry and geometric analysis*, volume 42005. Springer.
- Karatzas, I. & Shreve, S. (2012). *Brownian motion and stochastic calculus*. Springer Science & Business Media.
- Lebedev, L. P. & Cloud, M. J. (2003). *Tensor analysis*. World Scientific.
- Lee, J. M. (2002). *Introduction to smooth manifolds*. Springer.
- Leippold, M. & Wu, L. (2002). Asset pricing under the quadratic class. *Journal of Financial and Quantitative Analysis*, 37(2), 271–295.
- Leone, F., Nelson, L., & Nottingham, R. (1961). The folded normal distribution. *Technometrics*, 3(4), 543–550.
- Li, H. & Zhao, F. (2006). Unspanned stochastic volatility: evidence from hedging interest rate derivatives. *Journal of Finance*, 61(1), 341–378.
- Li, H. & Zhao, F. (2009). Nonparametric estimation of state-price densities implicit in interest rate cap prices. *Review of Financial Studies*, 22(11), 4335–4376.
- Litterman, R. B. & Scheinkman, J. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 1(1), 54–61.
- Longstaff, F. A., Santa-Clara, P., & Schwartz, E. S. (2001). The relative valuation of caps and swaptions: theory and empirical evidence. *Journal of Finance*, 56(6), 2067–2109.
- Lord, R., Koekkoek, R., & Dijk, D. V. (2010). A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2), 177–194.
- Meyer, C. D. (2000). *Matrix analysis and applied linear algebra*. Siam.
- Murray, M. P. (1994). A drunk and her dog: an illustration of cointegration and error correction. *The American Statistician*, 48(1), 37–39.
- Peaceman, D. W. & Rachford, H. H. (1955). The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics*, 3(1), 28–41.
- Piazzesi, M. (2010). Affine term structure models. *Handbook of Financial Econometrics*, 1, 691–766.

- Rady, S. (1997). Option pricing in the presence of natural boundaries and a quadratic diffusion term. *Finance and Stochastics*, 1(4), 331–344.
- Ritchken, P. & Sankarasubramanian, L. (1995). Volatility structures of forward rates and the dynamics of the term structure. *Mathematical Finance*, 5(1), 55–72.
- Santa-Clara, P. & Sornette, D. (2001). The dynamics of the forward interest rate curve with stochastic string shocks. *Review of Financial Studies*, 14(1), 149–185.
- Selby, M. J. & Strickland, C. (1995). Computing the Fong and Vasicek pure discount bond price formula. *Journal of Fixed Income*, 5(2), 78–84.
- Shreve, S. E. (2004). *Stochastic calculus for finance II: continuous-time models*. Springer Science & Business Media.
- Strang, G. (1993). The fundamental theorem of linear algebra. *The American Mathematical Monthly*, 100(9), 848–855.
- Teschl, G. (2012). *Ordinary differential equations and dynamical systems*. American Mathematical Society.
- Trolle, A. B. & Schwartz, E. S. (2009). A general stochastic volatility model for the pricing of interest rate derivatives. *Review of Financial Studies*, 22(5), 2007–2057.
- Vasiček, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), 177–188.